

EXPRESS MAIL NUMBER  
EL852296784 US

5

**SEQUENCE SPECIFIC RECOMBINASE-BASED METHODS FOR  
PRODUCING INTRON CONTAINING VECTORS AND COMPOSITIONS FOR  
USE IN PRACTICING THE SAME**

10

**CROSS-REFERENCE TO RELATED APPLICATIONS**

Pursuant to 35 U.S.C. §119(e), this application claims priority to the filing date of United States Provisional Patent Application Serial No. 60/263,358 filed January 18, 2001; the disclosure of which applications is herein incorporated by reference.

15

**INTRODUCTION**

**Field of the Invention**

The field of this invention is molecular biology, particularly recombinant DNA engineering.

**Background of the Invention**

The processes of isolating, cloning and expressing genes are central to the field of molecular biology and play prominent roles in research and industry in biotechnology and related fields. Until recently, the isolation and cloning of genes has been achieved *in vitro* using restriction endonucleases and DNA ligases. Restriction endonucleases are enzymes which recognize and cleave double-stranded DNA at a specific nucleotide sequence, and DNA ligases are enzymes which join fragments of DNA together via the phosphodiester bond. A DNA sequence of interest can be "cut" or digested into manageable pieces using a restriction endonuclease and then inserted into an appropriate vector for cloning using DNA ligase. However, in order to transfer the DNA of interest into a

different vector--most often a specialized expression vector--restriction enzymes must be used again to excise the DNA of interest from the cloning vector, and then DNA ligase is used again to ligate the DNA of interest into the chosen expression vector.

5       The ability to transfer a DNA of interest to an appropriate expression vector is often limited by the availability or suitability of restriction enzyme recognition sites. Often multiple restriction enzymes must be employed to remove the desired coding region. Further, the reaction conditions used for each enzyme may differ such that it is necessary to perform the excision reaction in separate  
10 steps, or it may be necessary to remove a particular enzyme used in an initial restriction enzyme reaction prior to completing subsequent restriction enzyme digestions due to buffer and/or cofactor incompatibility. Many of these extra steps require time-consuming purification of the subcloning intermediate.

15      There is, therefore, a need to develop protocols and compositions for the rapid transfer of a DNA molecule of interest from one vector to another *in vitro* or *in vivo* without the need to rely upon restriction enzyme digestions. To address this need, a number of different sequence specific recombinase based methods have been developed which allow one to transfer sequence material among vectors without restriction enzyme digestions. These systems include the  
20 commercially available Creator and Gateway sequence specific recombinase based methods, where representative systems are described in U.S. Patent Nos. 5,581,808 and 5,888,732; as well as in Published PCT Application Serial Nos. WO 00/12687 and WO 01/05961.

25      While the above protocols and systems are effective, there is room for improvement. For example, in the above systems, expression vectors that are produced by the methods encode fusion proteins of the gene of interest fused to a sequence encoded by the sequence specific recombinase site of the vector. In many instances, such a fusion sequence is undesirable.

30      As such, there is continued interest in the improvement of these sequence specific recombinase systems. Of particular interest would be the development of such a system that produced expression vectors where the protein of interest was

not expressed a fusion with sequence specific recombinase encoded sequences. The present invention satisfies this interest.

#### Relevant Literature

5 References of interest include: U.S. Patent Nos. 5,527,695; 5,744,336; 5,851,808; 5,888,732; and 5,962,255; as well as in Published PCT Application Serial Nos. WO 00/12687 and WO 01/05961. Also of interest is: Kaartinen & Nagy, Genesis (2001) 31: 126-129; and Yoshimura et al., Mol. Urol. (2001) 5: 81-4.

10

#### SUMMARY OF THE INVENTION

Methods are provided for producing a vector that includes at least one splicable intron. In the subject methods, intron containing vectors are produced from donor and acceptor vectors that each include a sequence specific

15 recombinase site, where the subject donor and acceptor vectors further include splice donor and acceptor sites that, upon sequence specific recombination of the donor and acceptor vectors, define an intron in the product vector of the recombination step. Also provided are compositions for use in practicing the subject methods, including the donor and acceptor vectors themselves, as well as  
20 systems and kits that include the same. The subject invention finds use in a variety of different applications, including the production of expression vectors that encode C-terminal tagged fusion proteins, the production of expression vectors that encode pure protein and not a fusion thereof with N- and/or C-terminal sequence specific recombinase site encoded residues, and the like.

25

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1 provides a map of the pDNR-Dual donor vector described in greater detail below.

30 Figure 2 provides a map of the pLPS-EGFP acceptor vector described in greater detail below.

Figure 3 provides a map of the pDNR-Dual-Luc vector described in greater detail below.

Figure 4 provides a map of the pLPS-Luc-EGFP vector described in greater detail below.

5       Figure 5 provides a flow diagram of a representative method according to the subject invention.

## DEFINITIONS

10       The terms "sequence-specific recombinase" and "site-specific recombinase" refer to enzymes or recombinases that recognize and bind to a short nucleic acid site or "sequence-specific recombinase target site", i.e., a recombinase recognition site, and catalyze the recombination of nucleic acid in relation to these sites. These enzymes include recombinases, transposases and  
15       integrases.

      The terms "sequence-specific recombinase target site", "site-specific recombinase target site", "sequence-specific target site" and "site-specific target site" refer to short nucleic acid sites or sequences, i.e., recombinase recognition sites, which are recognized by a sequence- or site-specific recombinase and  
20       which become the crossover regions during a site-specific recombination event. Examples of sequence-specific recombinase target sites include, but are not limited to, lox sites, att sites, dif sites and frt sites.

      The term "lox site" as used herein refers to a nucleotide sequence at which the product of the cre gene of bacteriophage P1, the Cre recombinase, can  
25       catalyze a site-specific recombination event. A variety of lox sites are known in the art, including the naturally occurring loxP, loxB, loxL and loxR, as well as a number of mutant, or variant, lox sites, such as loxP511, loxP514, loxΔ86, loxΔ117, loxC2, loxP2, loxP3 and lox P23.

      The term "frt site" as used herein refers to a nucleotide sequence at which  
30       the product of the FLP gene of the yeast 2 micron plasmid, FLP recombinase, can catalyze site-specific recombination.

The term "unique restriction enzyme site" indicates that the recognition sequence of a given restriction enzyme appears once within a nucleic acid molecule.

A restriction enzyme site or restriction site is said to be located "adjacent to

5 the 3' end of a sequence-specific recombinase target site" if the restriction enzyme recognition site is located downstream of the 3' end of the sequence-specific recombinase target site. The adjacent restriction enzyme site may, but need not, be contiguous with the last or 3' most nucleotide comprising the sequence-specific recombinase target site.

10 The term "intron" as used herein refers to a domain of a vector produced by the subject methods that is flanked on the 5' end by a splice donor site and on the 3' end by a splice acceptor site, where under appropriate conditions the intron is spliced out of or removed from an mRNA sequence expressed from the vector in which it is present.

15 The term "splice donor site" as used herein refers to a sequence or domain of a nucleic acid present at the 5' end of an intron, as defined above, that marks the start of the intron and its boundary with the preceding coding sequence – exon.

20 The term "splice acceptor site" as used herein refers to a sequence or domain of a nucleic acid present at the 3' end of an intron, as defined above, that marks the start of the intron and its boundary with the following coding sequence – exon.. In the present invention, the splice acceptor site is also meant to include the intron Branch point, which is required together with the splice donor and splice acceptor sequence in order for splicing to occur. The branch point marks

25 the point to which the 5'end of the intron becomes joined during the process of splicing. For convenience, in the present embodiments, the splice Acceptor sequence and the Branch site are placed adjacent to each other so that they can be encoded within a single synthetic oligonucleotide for ease of vector construction. Thus, they are described here as a single unit. However, they may

30 be further separated, by moving the branch site further 5' of the splice acceptor

sequence, provided that it is not moved 5' of the splice donor sequence and provided that splicing efficiency is not hindered.

The Term "splice site" as used herein refers to a sequence or domain of a nucleic acid present at either the 5' end or the 3' end of an intron as defined above.

The terms "polylinker" or "multiple cloning site" refer to a cluster of restriction enzyme sites, typically unique sites, on a nucleic acid construct that can be utilized for the insertion and/or excision of nucleic acid sequences, such as the coding region of a gene, loxP sites, etc.

The term "termination sequence" refers to a nucleic acid sequence which is recognized by the polymerase of a host cell and results in the termination of transcription. Prokaryotic termination sequences commonly comprise a GC-rich region that has a two-fold symmetry followed by an AT-rich sequence. A commonly used termination sequence is the T7 termination sequence. A variety of termination sequences are known in the art and may be employed in the nucleic acid constructs of the present invention, including the TINT3, TL13, TL2, TR1, TR2, and T6S termination signals derived from the bacteriophage lambda, and termination signals derived from bacterial genes, such as the trp gene of *E. coli*.

The terms "polyadenylation sequence" (also referred to as a "poly A<sup>+</sup> site" or "poly A<sup>+</sup> sequence") as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable, as transcripts lacking a poly A<sup>+</sup> tail are typically unstable and rapidly degraded. The poly A<sup>+</sup> signal utilized in an expression vector may be "heterologous" or "endogenous". An endogenous poly A<sup>+</sup> signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly A<sup>+</sup> signal is one which is isolated from one gene and placed 3' of another gene, e.g., coding sequence for a protein. A commonly used heterologous poly A<sup>+</sup> signal is the SV40 poly A<sup>+</sup> signal. The SV40 poly A<sup>+</sup> signal is contained on a 237 bp *Bam*H*I/Bc*I restriction fragment and directs both termination and polyadenylation;

numerous vectors contain the SV40 poly A<sup>+</sup> signal. Another commonly used heterologous poly A<sup>+</sup> signal is derived from the bovine growth hormone (BGH) gene; the BGH poly A<sup>+</sup> signal is also available on a number of commercially available vectors. The poly A<sup>+</sup> signal from the Herpes simplex virus thymidine kinase (HSV tk) gene is also used as a poly A<sup>+</sup> signal on a number of commercial expression vectors.

As used herein, the terms "selectable marker" or "selectable marker gene" refer to a gene which encodes an enzymatic activity and confers the ability to grow in medium lacking what would otherwise be an essential nutrient; in addition, a selectable marker may confer upon the cell in which the selectable marker is expressed, resistance to an antibiotic or drug. A selectable marker may be used to confer a particular phenotype upon a host cell. When a host cell must express a selectable marker to grow in selective medium, the marker is said to be a positive selectable marker (e.g., antibiotic resistance genes which confer the ability to grow in the presence of the appropriate antibiotic). Selectable markers can also be used to select against host cells containing a particular gene; selectable markers used in this manner are referred to as negative selectable markers.

As used herein, the term "construct" is used in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "vector" is sometimes used interchangeably with "construct". The term "construct" includes circular nucleic acid constructs such as plasmid constructs, phagemid constructs, cosmid vectors, etc., as well as linear nucleic acid constructs including, but not limited to, PCR products. The nucleic acid construct may comprise expression signals such as a promoter and/or an enhancer in operable linkage, and then is generally referred to as an "expression vector" or "expression construct".

The term "expression construct" as used herein refers to an expression module or expression cassette made up of a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a

particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

5       The terms "in operable combination", "in operable order" and "operably linked" as used herein refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The terms also refer to the linkage of amino acid sequences in such a manner so that  
10      the reading frame is maintained and a functional protein is produced.

A cell has been "transformed" or "transfected" with exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the  
15      transforming DNA may be maintained on an episomal element such as a vector or plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells  
20      containing the transforming DNA. A "clone" is a population of cells derived from a single cell or ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations. An organism, such as a plant or animal, that has been transformed with exogenous DNA is termed "transgenic".

25       Transformation of prokaryotic cells may be accomplished by a variety of means known in the art, including the treatment of host cells with CaCl<sub>2</sub> to make competent cells, electroporation, etc. Transfection of eukaryotic cells may be accomplished by a variety of means known in the art, including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection,  
30      polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, and biolistics.

As used herein, the term "host" is meant to include not only prokaryotes, but also eukaryotes, such as yeast, plant and animal cells. A recombinant DNA molecule or gene can be used to transform a host using any of the techniques commonly known to those of ordinary skill in the art. Prokaryotic hosts may 5 include *E. coli*, *S. typhimurium*, *Serratia marcescens* and *Bacillus subtilis*.

Eukaryotic hosts include yeasts such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, mammalian cells and insect cells, and, plant cells, such as *Arabidopsis thaliana* and *Tobaccum nicotiana*.

As used herein, the terms "restriction endonucleases" and "restriction 10 enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

"Recombinant DNA technology" refers to techniques for uniting two heterologous DNA molecules, usually as a result of *in vitro* ligation of DNAs from different organisms. Recombinant DNA molecules are commonly produced by 15 experiments in genetic engineering. Synonymous terms include "gene splicing", "molecular cloning" and "genetic engineering". The product of these manipulations results in a "recombinant" or "recombinant molecule". The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule that is expressed from a recombinant DNA molecule.

The ribose sugar is a polar molecule, and therefore, DNA is referred to as having a 5' to 3', or 5' to 3', directionality. DNA is said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor via a phosphodiester linkage. Therefore, 20 an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also has a 5' to 3' orientation. In either a linear or circular DNA 25 molecule, discrete elements are referred to as being "upstream" or "5'" of the "downstream" or "3'" elements. This terminology reflects the fact that DNA has an 30

inherent 5' to 3' polarity, and transcription typically proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements which direct transcription of an operably linked coding region, or open reading frame, are generally located 5', or upstream, of the coding region. However, enhancer  
5 elements can exert their effect even when located 3' of the promoter and coding region. Transcription termination and polyadenylation signals are typically located 3' or downstream of the coding region.

The 3' end of a promoter is said to be located upstream of the 5' end of a sequence-specific recombinase target site when, moving in a 5' to 3' direction  
10 along the nucleic acid molecule, the 3' terminus of a promoter precedes the 5' end of the sequence-specific recombinase target site. When the acceptor construct is intended to permit the expression of a translation fusion, the 3' end of the promoter is located upstream of both the sequences encoding the amino-terminus of a fusion protein and the 5' end of the sequence-specific recombinase  
15 target site. Thus, the sequence-specific recombinase target site is located within the coding region of the fusion protein (*i.e.*, located downstream of both the promoter and the sequences encoding the affinity domain, such as Gst).

As used herein, the term "adjacent", in the context of positioning of genetic elements in the constructs, shall mean within about 0 to 2500, sometimes 0 to  
20 1000 bp and sometimes within about 0 to 500, 0 to 400, 0 to 300 or 0 to 200 bp.

A DNA "coding sequence" is a double-stranded DNA sequence that is transcribed and translated into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a  
25 translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (*e.g.*, mammalian) DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence. A "cDNA"  
30 is defined as copy-DNA or complementary-DNA, and is a product of a reverse transcription reaction from an mRNA transcript. An "exon" is an expressed

sequence transcribed from the gene locus, whereas an "intron" is a non-expressed sequence that is from the gene locus.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell. A "cis-element" is a nucleotide sequence, also termed a "consensus sequence" or "motif," that interacts with proteins that can upregulate or downregulate expression of a specific gene locus. A "signal sequence" can also be included with the coding sequence. This sequence encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell and directs the polypeptide to the appropriate cellular location. Signal sequences can be found associated with a variety of proteins native to prokaryotes and eukaryotes.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction)

coding sequence. For purposes of defining the present invention, the promoter sequence includes, at its 3' terminus, the transcription initiation site and extends upstream (in the 5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters often, but not always, contain "TATA" boxes and "CAT" boxes.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length.

As used herein, "an origin of replication" or "origin" refers to any sequence capable of directing replication of a DNA construct in a suitable prokaryotic or eukaryotic host (e.g., the ColE1 origin and its derivatives; the yeast 2  $\mu$  origin). Eukaryotic expression vectors may also contain "viral replicons" or "origins of

replication". Viral replicons are viral DNA sequences which allow for the extrachromosomal replication of a vector in a host cell expressing the appropriate replication factors. Vectors which contain either the SV40 or polyoma virus origin of replication replicate to high copy number (up to  $10^4$  copies/cell) in cells that 5 express the appropriate viral T antigen. Vectors which contain the replicons from bovine papillomavirus or Epstein-Barr virus replicate extrachromosomally at low copy number (~100 copies/cell).

As used herein, the terms "nucleic acid molecule encoding", "DNA sequence encoding", and "DNA encoding" refer to the order or sequence of 10 deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

As used herein, the term "gene" means the deoxyribonucleotide sequences comprising the coding region of a structural gene, i.e., the coding 15 sequence for a protein or polypeptide of interest, including sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end, such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The 20 sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences". 25 Introns are segments of a gene that are transcribed into heteronuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the mature messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino 30 acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated

5 sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

10 As used herein, the term "purified" or "to purify" refers to the removal of contaminants from a sample. For example, recombinant Cre polypeptides are expressed in bacterial host cells (e.g., as a GST-Cre or (HN)<sub>6</sub>-Cre fusion protein) and the Cre polypeptides are purified by the removal of host cell proteins; the percent of recombinant Cre polypeptides is thereby enriched or increased in the

15 sample.

As used herein the term "portion" refers to a fraction of a sequence, gene or protein. "Portion" may comprise a fraction greater than half of the sequence, gene or protein, equal to half of the sequence, gene or protein or less than half of the sequence, gene or protein. Typically as used herein, two or more "portions"

20 combine to comprise a whole sequence, gene or protein.

As used herein, the term "fusion protein" refers to a chimeric protein containing a protein of interest joined to an exogenous protein fragment. The fusion partner may enhance solubility of the protein of interest as expressed in a host cell, may provide an affinity tag to allow purification of the recombinant fusion

25 protein from the host cell or culture supernatant, or both. If desired, the fusion protein may be removed from the protein of interest by a variety of enzymatic or chemical means known to the art.

#### DESCRIPTION OF THE SPECIFIC EMBODIMENTS

30 Methods are provided for producing a vector that includes at least one splicable intron. In the subject methods, intron containing vectors are produced

from donor and acceptor vectors that each include a site specific recombinase site, where the subject donor and acceptor vectors further include splice donor and acceptor sites that, upon site specific recombination of the donor and acceptor vectors, define an intron in the product vector of the recombination step.

5 Also provided are compositions for use in practicing the subject methods, including the donor and acceptor vectors themselves, as well as systems and kits that include the same. The subject invention finds use in a variety of different applications, including the production of expression vectors that encode C-terminal tagged fusion proteins, the production of expression vectors that encode

10 pure protein and not a fusion thereof, and the like.

Before the subject invention is described further, it is to be understood that the invention is not limited to the particular embodiments of the invention described below, as variations of the particular embodiments may be made and 15 still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the purpose of describing particular embodiments, and is not intended to be limiting. Instead, the scope of the present invention will be established by the appended claims.

20 In this specification and the appended claims, the singular forms "a," "an" and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

25 Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range, and any other stated or intervening value in that stated range, is encompassed within the invention.

30 The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention,

subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

5        Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now  
10      described.

All publications mentioned herein are incorporated herein by reference for the purpose of describing various invention components that are described in the publications which might be used in connection with the presently described  
15      invention.

In further describing the subject invention, the subject methods are reviewed first in greater detail, followed by a review of representative applications in which the subject methods find use, as well as a review of systems, libraries  
20      and kits for use in practicing the subject methods.

## METHODS

As summarized above, the subject invention provides recombinase-based  
25      methods for producing intron containing vectors. In other words, the subject invention provides methods of producing vectors that include at least one intron, where the methods are site specific recombinase based methods. By “site specific recombinase” based method is meant that the subject methods employ a recombinase mechanism to produce the subject intron containing vectors. The  
30      recombinase mechanism that is employed in the subject methods is one in which a recombinase mediates the transfer of a nucleic acid from a donor to an

acceptor vector, where the donor and acceptor vectors each include at least one recombinase recognition site. A variety of different site specific recombinase systems suitable for transferring a nucleic acid from a donor to an acceptor vector are known and may be modified to be useful in the subject invention. Such

5 systems include those described in U.S. Patent Nos. 5,851,808; 5,888,732; and U.S. Provisional Application Serial No. 09/616,651, the disclosure of which are herein incorporated by reference, as well as WO 00/12687 and WO 01/05961, the disclosures of the priority documents of which are herein incorporated by reference.

10 In general, in addition to each including at least one recombinase recognition site, the donor and acceptor vectors each include at least one splice site, e.g., a splice donor site or a splice acceptor site. In certain embodiments, the donor and acceptor vectors each include a single splice site, where in many of these embodiments, the donor vector includes a splice donor site and the

15 acceptor vector includes a splice acceptor site. In yet other embodiments, the donor and acceptor vectors each include splice donor and acceptor sites which are oriented such that they do not form an intron in the donor vectors but, upon recombinase mediated recombination of the donor and acceptor vectors, produce a resultant vector with two distinct introns. In such designs, the acceptors will

20 contain one synthetic intron that encompasses the recombinase recognition sequence and the acceptor partial selectable marker.

Any convenient splice sites (i.e., splice donor and acceptor sites) may be employed in the vectors of the subject method. Representative splice sites or sequences, e.g., domains, of interest that may be employed include both splice sites that require specifically provided factors for splicing, e.g., eukaryotic host factors (as found in a eukaryotic host cells) such that the intron is only spliced in a eukaryotic host cell or a mimetic (e.g., *in vivo* or *in vitro*) environment that provides all the relevant factors, and splice sites that are self-splicing or autocatalytic, i.e., do not require specific factors for splicing to occur, and thus are

25 spliced in both eukaryotic and prokaryotic environments, as well as *in vitro* environments. Examples include the splicing elements of Group I and Group II

self-splicing introns found in bacteria, and certain cellular organelles, e.g., the highly conserved in Group I self-splicing intron, P7; the bacterial group II intron *L. lactis* L1.ltrB; the yeast mitochondrial group II introns al1 and al2; and the bacterial group II intron *Sinorhizobium meliloti* RmIInt1 (see Oe Y., et al., 2001; and  
5 Martinez-Abarca, F. and Toro, N., 2000)

Any convenient splice acceptor donor and acceptor sites may be employed. Consensus sequences for the 5' splice donor site and the 3' splice acceptor site used in RNA splicing are well known in the art (See, Moore, et al.,  
10 1993, *The RNA World*, Cold Spring Harbor Laboratory Press, p. 303-358). In addition, modified consensus sequences that maintain the ability to function as 5' donor splice sites and 3' splice acceptors sites may be used in the practice of the invention. In certain embodiments, splice-donor sites have a characteristic consensus sequence represented as: (A/C)AGGURAGU (where R denotes a  
15 purine nucleotide) with the GU in the fourth and fifth positions being required (Jackson, I. J., *Nucleic Acids Research* 19: 3715-3798 (1991)). Splice-donor sites are functionally defined by their ability to effect the appropriate reaction within the mRNA splicing pathway. An unpaired splice-donor site is defined herein as a  
20 splice-donor site which is present in a donor or acceptor vector, typically a donor vector, and is not accompanied in the vector by a splice-acceptor site positioned 3' to the unpaired splice-donor site. Upon recombinase mediated recombination between the donor and acceptor vectors, the unpaired splice-donor site results in splicing to a splice-acceptor site originally present in the other vector. A splice-acceptor site is a sequence which, like a splice-donor site, directs the splicing of  
25 an intron out of a resultant expression cassette produced upon recombinase mediated recombination of the donor and acceptor vectors. Acting in conjunction with a splice-donor site, the splicing apparatus uses a splice-acceptor site to effect the removal of an intron. Splice-acceptor sites have a characteristic sequence represented as: YYYYYYYYYYYNYAG, where Y denotes any pyrimidine  
30 and N denotes any nucleotide (Jackson, I. J., *Nucleic Acids Research* 19:3715-3798 (1991)). For convenience, in the present embodiments, the splice acceptor

sequence is immediately preceded by the intron Branch site and these are considered here as one unit, although they may be separated. The consensus Branch site is: YNYYRAY, where Y denotes any pyrimidine, R any purine, and N denotes any nucleotide.

5 Specific splice sites of interest include, but are not limited to: (a) the novel consensus intron sequences and the Human hemoglobin Beta donor and acceptor sequences described in Liu Z. et al Anal Biochem 246: 264-267 (1997) and found in the experimental section, infra; (b) the donor and acceptor sequences found in the SV40 late 19s and 16s mRNA introns (see pCMV myc  
10 from Clontech ); (c) the splice donor and acceptor sequences found in the rabbit Beta globin intron (found in the vector pCMV-neo-Bam); and the like.

The position of the splice donor and acceptor sequences in the various donor and acceptor vectors determines the location of the intron in the resultant product vector and, therefore, the domain that is spliced out of the resultant  
15 vector under appropriate splicing conditions, e.g., in a eukaryotic host cell. Thus, by knowing how the acceptor and donor vectors recombine into a resultant vector, one can position the donor and acceptor splice sites in the donor and acceptor vectors to provide for an intron in any location of the resultant vector, and therefore removal of any sequence of the resultant vector. For example, the  
20 donor and acceptor splice sites can be positioned to provide for a spliceable intron in the resultant product vector that includes the 3' recombinase recognized site, the 5' recombinase recognized site, etc. See, e.g., the experimental section below for more details with respect to a donor and acceptor vector system in which the donor and acceptor splice sites are positioned to provide for a resultant  
25 vector in which the 3' recombinase site (lox) is present in a spliceable intron.

In many embodiments of interest, the donor and acceptor vectors are further characterized in that one of the donor and acceptor vectors includes only one recombinase recognition site, while the other of the donor and acceptor vectors includes two recombinase recognition sites. As mentioned above, in many  
30 embodiments, the donor vector includes two recombinase recognition sites while the acceptor vector includes a single recombinase recognition site. In an

alternative embodiment, the donor vector includes a single recombinase recognition site while the acceptor vector includes two recombinase recognition sites. Such a system is described in U.S. Application Serial No. 09/616,651, the disclosure of which is herein incorporated by reference.

5 A feature of the vectors of these embodiments is that the donor and acceptor vectors must be able to recombine in the presence of a suitable recombinase to produce an expression vector as described above, where the expression vector lacks at least a portion of the initial donor or acceptor vector, i.e., it is a non-fusion expression vector. As such, the donor and acceptor vectors  
10 must be able to participate in a recombination event that is other than a fusion event, where by fusion event is meant an event in which two complete vectors are fused in their entirety into one fused vector, e.g., where two plasmids are fused together to produce one plasmid that includes all of material from the initial two plasmids, i.e., a fusion plasmid. As such, the subject methods of these particular  
15 embodiments are not fusion methods, where such methods are defined as those methods in which a single vector is produced from two or more initial vectors in their entirety, such that all of the initial vector material of each parent vector, e.g., plasmid, is present in its entirety in the resultant fusion vector.

The donor and acceptor vectors of these particular embodiments are  
20 further characterized in that one of the donor and acceptor vectors includes only one recombinase recognition site, while the other of the donor and acceptor vectors includes two recombinase recognition sites. In a first preferred embodiment, the donor vector includes two recombinase recognition sites while the acceptor vector includes a single recombinase recognition site. In an  
25 alternative embodiment, the donor vector includes a single recombinase recognition site while the acceptor vector includes two recombinase recognition sites. The donor and acceptor vectors of this first, preferred embodiment and this second, alternative embodiment, are described in greater detail below.

The donor and acceptor vectors described generally above may be linear  
30 or circular, e.g., plasmids, and in many embodiments of the subject invention are plasmids. Where the donor and acceptor vectors are plasmids, the donor and

acceptor vectors typically range in length from about 2 kb to 200 kb, usually from about 2 kb to 40 kb and more usually from about 2 kb to 10 kb.

The donor and acceptor vectors are further characterized in certain embodiments in that all of the recombinase recognition sites on the donor and

5 acceptor vectors must be recognized by the same recombinase and should be able to recombine with each other, but within this parameter they may be the same or different, but in many embodiments are usually the same. Recombinase recognition sites, i.e., sequence-specific recombinase target sites, of interest include: Cre recombinase activity recognized sites, e.g., loxP, loxP2, loxP511,  
10 loxP514, loxB, loxC2, loxL, loxR, loxΔ86, loxΔ117; att, dif; frt; and the like. The particular recombinase recognition site is chosen, at least in part, based on the nature of the recombinase to be employed in the subject methods.

#### *The Donor Vector*

15 As mentioned above, in a preferred embodiment of the subject methods, the donor vector includes two recombinase recognition sites while the acceptor vector includes a single recombinase recognition site. In the donor vector of these embodiments, the donor vector includes two recombinase recognition sites,  
20 capable of recombining with each other, e.g., site 1A and site 1B, that flank or border a first or donor domain, i.e., desired donor fragment, where this domain is the portion of the vector that becomes part of the expression vector produced by the subject methods. The length of the donor domain may vary, but in many embodiments ranges from 1 kb to 200 kb, usually from about 1 kb to 10 kb. The  
25 portion of the donor vector that is not part of this donor domain, i.e., the part that is 5' of site 1A and 3' of site 1B, is referred to herein for clarity as the non-donor domain of the donor vector.

The two recombinase recognition sites of the donor vector are characterized in that they are oriented in the same direction and are capable of recombining with each other. By oriented in the same direction it is meant that

they have the same head to tail orientation. Thus, the orientation of site 1A is the same as the orientation of site 1B.

The donor domain flanked by the two recombinase recognition sites, i.e., the portion of the vector 3' of the first recombinase site 1A and 5' of the second recombinase site 1B, includes at least the following components: (a) at least one restriction site and (b) at least a portion of a selectable marker, e.g. a coding sequence, a promoter, or a complete selectable marker made up of a coding sequence and a promoter. The donor domain may include at least one restriction site or a plurality of distinct restriction sites, e.g., as found in a multiple cloning site or polylinker, where by restriction site is meant a stretch of nucleotides that has a sequence that is recognized and cleaved by a restriction endonuclease. Where a plurality of restriction sites are present in the donor domain, the number of distinct or different restriction sites typically ranges from about 2 to 5, usually from about 2 to 13.

In many embodiments, there are at least two restriction sites, which may or may not be identical depending on the particular protocol employed to produce the donor plasmid, that flank a nucleic acid which is a coding sequence for a protein of interest, where the protein of interest may or may not be known, e.g., it may be a known coding sequence for a known protein or polypeptide or a coding sequence for an as yet unidentified protein or polypeptide, such as where this nucleic acid of interest is a constituent of a library, as discussed in greater detail below. The length of this nucleic acid of interest nucleic acid may vary greatly, but generally ranges from about 18 bp to 20 kb, usually from about 100 bp to 10 kb and more usually from about 1 kb to 3 kb. At least one restriction site and this nucleic acid of interest nucleic acid, when present, are sufficiently close to the 3' end of the first flanking recombinase site, i.e., recombinase recognition site 1A, such that in the expression vector produced from the donor plasmid, expression of the coding sequence of the nucleic acid of interest is driven by a promoter positioned 5' of this first recombinase site. As such, the distance separating this restriction site/nucleic acid of interest nucleic acid from the recombinase site typically ranges from about 1 bp to 150 bp, usually from about 1 bp to 50 bp.

In a first preferred embodiment, the donor domain also generally includes a portion of a selectable marker. By portion of a selectable marker is meant a sub-part of a selectable marker, e.g. a coding sequence or a promoter, which can be joined with a second subpart to produce a functioning selectable marker that

5 confers some selectable phenotype on the host cell in which the expression vector produced by the subject methods is to be propagated. Examples of subparts of selectable markers are coding sequences and promoters. As such, in many embodiments, the portion of the selectable marker present on the donor domain is a coding sequence of a marker gene or a promoter capable of driving

10 expression of the coding sequence of the marker gene, where in certain preferred embodiments, the coding sequence of a marker gene is the portion of the selectable marker present on the donor domain. Examples of coding sequences of interest include, but are not limited to, the coding sequences from the following marker genes: the chloramphenicol resistance gene, the ampicillin resistance

15 gene, the tetracycline resistance gene, the kanamycin resistance gene, the streptomycin resistance gene and the SacB gene from *B. subtilis* encoding sucrase and conferring sucrose sensitivity; and the like. The promoter portions or sub-parts of this selectable marker are any convenient promoters capable of driving expression of the selectable marker in the expression vector produced by

20 the subject methods, see *infra*, and in many embodiments are bacterial promoters, where particular promoters of interest include, but are not limited to: the Ampicillin resistance promoter, the inducible lac promoter, the tet-inducible promoter from pProTet ( $P_{tetO-1}$ )- available from CLONTECH, T7, T3, and SP6 promoters; and the like. The distance of this sub-part or portion of the selectable

25 marker from the 3' end of the second recombinase recognition site, i.e., site 1B, is sufficient to provide for expression of the marker to occur in the final expression vector, where the other part of selectable marker that is required for efficient expression of the selectable marker is present on the other side, i.e., the 5' side of the adjacent recombinase recognition site. This distance typically ranges from

30 about 1 bp to 2.5 kb, usually from about 1 bp to 500 bp.

The length of the donor domain flanked by the first and second recombinase sites of the donor plasmid, i.e., the length of the desired donor fragment, may vary greatly, so long as the above described components are present on the donor domain. Generally, the length is at least about 100 bp,

5 usually at least about 500 bp and more usually at least about 900 bp, where the length may be as great as 100 kb or greater, but generally does not exceed about 20 kb and usually does not exceed about 10 kb. Typically, the length of the donor domain ranges from about 100 bp to 100 kb, usually from about 500 bp to 20 kb and more usually from about 900 bp to 10 kb.

10 In addition to the above described components, the donor vector may include a number of additional elements, where desired, that are present on the non-donor domain or non-desired donor fragment of the donor vector. For example, the non-donor domain generally includes an origin of replication. This origin of replication may be any convenient origin of replication or ori site, where a

15 number of ori sites are known in the art, where particular sites of interest include, but are not limited to: ColE1 and its derivatives, pMB1, other origins that function in prokaryotic cells, the yeast 2 micron origin and the like. Also present on this non-donor domain of certain preferred embodiments is a selective marker gene that provides for negative selection of the non-donor domain under particular

20 conditions, e.g., negative selection conditions. This marker is fully functional and therefore is made up of a coding sequence operably linked to an appropriate promoter, i.e., is provided by a functional expression module or cassette. Markers of interest that are capable of providing for this negative selection include, but are not limited to: SacB, providing sensitivity to sucrose; ccdB; and the like.

25 This non-donor domain of the donor vector may further include one or more additional components or elements that impart additional functionality to the donor vector. For example, the donor vector may be a vector that is specifically designed for use in conjunction with a yeast two hybrid assay protocol, e.g., such that one can determine whether the gene of interest present in the donor domain

30 encodes a product that binds to a second protein prior to transferal of the gene of interest to an expression vector. In such embodiments, the non-donor domain

typically includes the following additional elements: yeast origins of replication, e.g., the yeast 2 micron origin; yeast selection markers, e.g., URA3, Leu, and trp selection markers; and peptide fragments of yeast transcription factors that are expressed as translational fusions to the gene encoded within the donor-domain;

5 where yeast two hybrid systems are known to those of skill in the art and described in: Fields, S. and O-K. Song. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**:245-246; Fields, S. and R. Sternglanz. 1994. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet.* **10**: 286-292 and the MATCHMAKER system III user manual, available

10 from CLONTECH.

In other embodiments, the non-donor domain and/or donor domains may contain yet other functional elements that provide specific functions to the donor. For example, Donor vectors can be designed that would also function as prokaryotic expression vectors that express the gene of interest encoded on the donor domain in prokaryotic cells either as a native protein or fused to an affinity or epitope tag. Such vectors may include the following elements in their non-donor or donor domains (e.g., 3' of the multiple cloning site): inducible bacterial promoters, such as the lac promoter or the  $P_{laco}$  promoter; affinity or epitope tags, e.g., GST, 6x(HN), myc-tag, HA-Tag, GFP and its derivatives. Donor

15 vectors designed to function as retroviral vectors would additionally include retroviral LTRs and packaging signals in the non-donor domain. Donor vectors for expression in mammalian cells might also encode affinity or epitope tags, e.g., GST, 6x(HN), myc-tag, HA-Tag, GFP and its derivatives; and mammalian constitutive or inducible promoters, e.g., the CMV promoter, the tet-inducible

20 promoter, the TK promoter; viral promoters, e.g., T7, T3, SP6. In a preferred embodiment of this particular embodiment of the subject invention, the donor vector is as follows. The donor-partial selectable marker comprises the open reading frame (ORF) for a selectable marker gene, and is placed between the two donor sequence-specific recombinase target sites, adjacent to the second-donor

25 sequence-specific recombinase target site. In a more preferred embodiment of the donor construct, the open reading frame of the selectable marker is situated

30

such that its 5' to 3' orientation is opposite that of the two donor sequence-specific recombinase target sites.

In another embodiment of the donor construct, the donor construct is a closed circle (e.g., a plasmid or cosmid) comprising, in addition to the two donor sequence-specific recombinase target sites, the unique restriction site or polylinker and the selectable marker gene open reading frame, at least one origin of replication, and at least one donor-functional selectable marker gene. The methods of the present invention should not be limited by the origin of replication selected. For example, origins such as those found in the pUC series of plasmid vectors or of the pBR322 plasmid may be used, as well as others known in the art. Those skilled in the art know that the choice of origin depends on the application for which the donor construct is intended and/or the host strain in which the construct is to be propagated.

A variety of selectable marker genes may be utilized, either for the donor-partial selectable marker or for the donor-functional selectable marker, and such genes may confer either positive- or negative-resistance phenotypes; however, the donor-partial and the donor-functional selectable marker genes should be different from one another. In a preferred embodiment, the selectable markers are selected from the group consisting of the chloramphenicol resistance gene, the ampicillin resistance gene, the tetracycline resistance gene, the kanamycin resistance gene, the streptomycin resistance gene and the *sacB* gene from *B. subtilis* encoding sucrase and conferring sucrose sensitivity. In a more preferred embodiment, the donor-partial selectable marker is a portion of the gene (e.g., the open reading frame) for chloramphenicol resistance and the donor-functional selectable marker gene is the gene for ampicillin resistance. In another preferred embodiment of the donor construct, the origin of replication and the donor-functional selectable marker gene lie 5' of the first-donor sequence-specific recombinase target site.

In another embodiment of the present invention, there is provided a donor construct with all the above-described features, but additionally having a marker gene different from either the donor-functional selectable marker gene or the

donor-partial selectable marker gene, wherein the additional marker gene is positioned 5' of the first sequence-specific recombinase target site such that upon combination with a recombinase, the additional marker gene is located on the undesired second donor fragment. This marker gene provides an additional

5 screen to exclude any products that result in recombinants containing the second donor fragment. The marker gene could be, for example, LacZ. In this case, incorrect recombinants would generate blue colonies on X-Gal plates.

Alternatively, a more preferred additional marker would be the sacB gene conferring sucrose sensitivity. In this case, any incorrect clones would be killed

10 when grown on sucrose containing medium. The additional marker provides another screen, thereby enhancing the system by further ensuring that only correct recombination products are obtained following recombination and transformation.

In yet another embodiment of the donor construct, the donor construct

15 further comprises a termination sequence placed 3' of the restriction site or polylinker sequence but 5' of the second-donor sequence-specific recombinase target site. In a most preferred embodiment, the termination sequence is placed 5' of the 3' end of the donor-partial selectable marker (e.g. the ORF of the selectable marker gene in the preferred embodiment which is in the 5' to 3'

20 orientation opposite that of both donor sequence specific recombinase target sites). The present embodiment is not be limited by the termination sequence chosen. In one embodiment, the termination sequence is the T1 termination sequence; however, a variety of termination sequences are known to the art and may be employed in the nucleic acid constructs of the present invention, including

25 the T6S, TINT, TL1, TL2, TR1, and TR2 termination signals derived from the bacteriophage lambda, and termination signals derived from bacterial genes such as the trp gene of *E. coli*.

In another preferred embodiment of the donor construct, the donor construct further comprises a polyadenylation sequence placed 3' of the unique

30 restriction site(s) or polylinker but 5' of the second-donor sequence-specific recombinase target site. In a most preferred embodiment, the polyadenylation

sequence is placed 5' of the 3' end of the open reading frame of the selectable marker gene similar to the placement described for the termination sequence *supra*. The present invention should not be limited by the nature of the polyadenylation sequence chosen. In one embodiment, the polyadenylation sequence is selected from the group consisting of the bovine growth hormone polyadenylation sequence, the simian virus 40 polyadenylation sequence and the Herpes simplex virus thymidine kinase polyadenylation sequence.

Also, in a preferred embodiment, the donor construct further comprises a gene or DNA sequence of interest inserted into the unique restriction enzyme site or polylinker. The present invention should not be limited by the size of the DNA of interest inserted into the unique restriction site or polylinker nor the source of DNA (e.g., genomic libraries, cDNA libraries, etc.).

Thus, in a most preferred embodiment of the donor nucleic acid construct, there is provided, in 5' to 3' order: a) a first-donor sequence-specific recombinase target site; b) a nucleic acid or gene of interest; c) termination and polyadenylation sequences; d) an open reading frame for a selectable marker gene in a 5' to 3' orientation opposite to that of the first-donor sequence-specific recombinase target site; e) a second-donor sequence-specific recombinase target site in the same 5' to 3' orientation as the first donor sequence-specific recombinase target site, wherein the second-donor sequence-specific recombinase target site is able to recombine with said first-donor sequence-specific recombinase target site; f) an origin of replication; and g) a donor-functional selectable marker gene.

In addition to the above features, the donor vector also includes at least one splice site, e.g., a splice donor and/or splice acceptor site. Two representative and non-limiting embodiments are now reviewed. In certain embodiments, the donor vector includes a splice donor site that is positioned to provide for an intron flanking the 3' sequence specific recombinase site in the product vector. In these embodiments, the splice donor site is positioned between the 5' and 3' sequence specific recombinase sites and, more usually, 3' of the multiple cloning site or gene of interest and 5' of the second sequence specific recombinase site. These

embodiments find use in producing vectors that express the gene of interest as a C-terminal tagged fusion, as a product that does not include sequence encoded by the 3' sequence specific recombinase site, etc. In certain embodiments, the donor vector also includes a splice acceptor site that is immediately 3' of the 5'

5 sequence specific recombinase site. Since the splice acceptor is 5' of the splice donor sites in the vector, the two splice sites do not make a spliceable intron in the donor vector. However, upon recombination with an appropriate acceptor vector, a product vector in which both the 5' and 3' sequence specific recombinase sites are present in distinct introns can be produced. These embodiments are useful in

10 applications where one wishes to express a protein from the product vector in a manner that is free of any residues encoded by the 5' and 3' sequence specific recombinase sites.

#### *The Acceptor Vector*

15 As mentioned above, in a preferred embodiment of the subject invention, the acceptor vector employed in the subject methods is a vector that includes a single recombinase site. In these embodiments, the single recombinase site is flanked on one side by a promoter and on the other side, in certain preferred

20 embodiments, by a portion of a selectable marker, e.g., a promoter or a coding sequence, where in many preferred embodiments described further below, this portion or sub-part of the selectable marker is a second promoter, e.g., a bacterial promoter. In these embodiments, the single recombinase site is flanked by two oppositely oriented promoters, where one of promoters drives expression of the

25 gene of interest in the expression vector produced by the subject methods and the second promoter drives expression of the coding sequence of the recombinant-functional selectable marker in the expression vector produced by the subject methods. In these embodiments, the first promoter is a promoter that is capable of driving expression of the gene of interest in the expression vector,

30 where representative promoters include, but are not limited to the CMV promoter, the tet-inducible promoter; retroviral LTR promoter/enhancer sequences, the TK

promoter, bacterial promoters, e.g. the lac promoter , the P<sub>LtetO-1</sub> promoter; the yeast ADH promoter and the like. The distance between the first promoter and the recombinase site is one that allows for expression in the final expression vector, where the distance typically ranges from about 1 bp to 1000 bp, usually from  
5 about 10 bp to 500 bp. The second promoter is a promoter that is capable of driving expression of the recombinant-functional selectable marker, and is generally a bacterial promoter. Bacterial promoters of interest include, but are not limited to: the Ampicillin promoter, the lac promoter , the P<sub>LtetO-1</sub> promoter , the T7 promoter and the like. The distance between the bacterial promoter and the  
10 recombinase site is sufficient to provide for expression of the selectable marker in the expression vector and typically ranges from about 1 bp to 2.5 kb, usually from about 1 bp to 200 bp.

As indicated above, in yet other preferred embodiments the acceptor vector lacks the portion or subpart of the selectable marker. In these  
15 embodiments, the acceptor vector may be used with a donor vector that includes a complete positive selectable marker in the desired donor fragment flanked by the two recombinase sites, i.e., the donor vector portion located between the 3' end of the first recombinase site and the 5' end of the second recombinase site. Alternatively, the acceptor vector may be used with a donor vector that only  
20 includes a partial selectable positive marker, as described above, where the partial marker is nonetheless functional in the resultant expression vector.

The acceptor vector of the embodiments described above may include a number of additional components or elements which are requisite or desired depending on the nature of the expression vector to be produced from the  
25 acceptor vector. In many embodiments of the subject invention, the acceptor vector is an acceptor nucleic acid construct comprising: a) an origin of replication capable of replicating the final desired recombination construct or expression vector; b) an acceptor sequence-specific recombinase target site having a defined 5' to 3' orientation; c) a first promoter adjacent to the 5' end of the  
30 acceptor sequence-specific recombinase target site; and d) an acceptor-partial selectable marker, wherein the acceptor-partial selectable marker is capable of

recombining with a donor-partial selectable marker from a donor construct (or first donor fragment, once the donor construct is resolved) so creating a recombinant-functional selectable marker in a final desired recombination construct. As in the donor construct, the acceptor construct is not limited by the nature of the

5 sequence-specific recombinase target site employed, and in preferred embodiments the sequence-specific recombinase target site may be selected from the group consisting of loxP, loxP2, loxP511, loxP514, loxB, loxC2, loxL, loxR, loxA86, loxA117, loxP3, loxP23, att, dif, and frt. The acceptor sequence-specific recombinase target site from the acceptor construct does not have to be

10 identical to those on the donor construct; however, the sequence-specific recombinase target sites on the acceptor and donor constructs must be able to recombine with each other.

In a preferred embodiment, the acceptor-partial selectable marker is a second promoter, wherein the second promoter is oriented such that its 5' to 3' orientation is opposite that of the acceptor sequence-specific recombinase target site and the first promoter, and wherein the 3' end of the second promoter is adjacent to the 3' end of the acceptor sequence-specific recombinase target site.

The acceptor construct is not limited by the nature of the origin of replication employed. A variety of origins of replication are known in the art and may be employed on the acceptor nucleic acid constructs of the present invention. Those skilled in the art know that the choice of origin depends on the application for which the acceptor construct is intended and/or the host strain in which the construct is to be propagated. In the case of the acceptor construct, the origin of replication is chosen appropriately such that both the acceptor construct and the final desired recombination construct will be able to replicate in the given host cell.

The acceptor construct also is not limited by the nature of the promoters employed. Those skilled in the art know that the choice of the promoter depends upon the type of host cell to be employed for expressing a gene(s) under the transcriptional control of the chosen promoter. A wide variety of promoters functional in viruses, prokaryotic cells and eukaryotic cells are known in the art

and may be employed in the acceptor nucleic acid constructs of the present invention. In a preferred embodiment of the invention, the donor construct contains a gene or DNA sequences of interest and when the donor construct recombines with the acceptor construct, the first promoter of the acceptor

5 construct is positioned such that it will drive expression of the gene or DNA sequences of interest. Thus, a promoter capable of driving the gene or DNA sequences of interest should be chosen for the first promoter. Further, in a preferred embodiment of the present invention, the acceptor-partial selectable marker is a promoter capable of driving the expression of the donor-partial  
10 selectable marker ORF from the donor construct (e.g., the promoter for the ampicillin gene from the plasmid pUC19) or a viral promoter including, but not limited to, the T7, T3, and Sp6 promoters.

In yet another preferred embodiment of the acceptor construct, the acceptor construct additionally includes a DNA sequence encoding a peptide

15 affinity domain or peptide tag sequence, wherein the affinity domain or tag sequence is 3' of the first promoter and 5' of the acceptor sequence-specific recombinase target site, such that the expression of the affinity domain or tag sequence is under control of the first promoter, and such that it is in the same translational frame as the acceptor sequence-specific recombinase target site.  
20 The present invention is not limited by the nature of the affinity domain or tag sequence employed; a variety of suitable affinity domains are known in the art, including glutathione-S-transferase, the maltose binding protein, protein A, protein L, polyhistidine tracts, etc.; and tag sequences include, but are not limited to the c-Myc Tag, the HA Tag, the FLAG tag, Green Fluorescent Protein (GFP), etc.

25 In another preferred embodiment of the acceptor vector construct, the acceptor construct additionally includes a DNA sequence encoding a peptide affinity domain or peptide tag sequence, wherein the affinity domain or tag sequence is 3' of an intron splice acceptor sequence placed in the acceptor vector 3' of the partial selectable marker, such that when this vector is  
30 recombined with a donor vector of the invention having an appropriately positioned intron splice donor sequence, an expression cassette is generated

having a functional synthetic intron and in which the expression of the affinity domain or tag sequence is under control of the first promoter of the acceptor vector, and such that it is in the same translational frame as a gene of interest placed within the donor vector. The present invention is not limited by the nature  
5 of the affinity domain or tag sequence employed; a variety of suitable affinity domains are known in the art, including glutathione-S-transferase, the maltose binding protein, protein A, protein L, polyhistidine tracts, etc.; and tag sequences include, but are not limited to the c-Myc Tag, the HA Tag, the FLAG tag, Green  
10 Fluorescent Protein (GFP), etc. Since this tag and the gene of interest are in-frame, following splicing, they will be expressed as a single fusion protein, with the Tag being at the C-terminus of the protein.

In another preferred embodiment of the acceptor construct, the acceptor construct further includes an acceptor-functional selectable marker. The present invention is not limited by the nature of the acceptor-functional selectable marker  
15 chosen and the selectable marker gene may result in positive or negative selection. In a preferred embodiment, the acceptor-functional selectable marker gene is selected from the group consisting of the chloramphenicol resistance gene, the ampicillin resistance gene, the tetracycline resistance gene, the kanamycin resistance gene, the streptomycin resistance gene and the sacB  
20 gene.

In addition to one or more of the above described components, the acceptor vectors may include a number of additional components that impart specific function to the expression vectors that are produced from the acceptor vector according to the subject methods. Additional elements that may be present  
25 on the subject acceptor vectors include, but are not limited to: (a) elements requisite for generating vectors suitable for use in yeast two hybrid expression assays, e.g., a GAL4 activation domain coding sequence, a GAL4 DNA-binding domain coding sequence, (as found in pLP-GADT7 and pLP-GBKT7 shown in Figs. 3A & 3B); (b) elements necessary for study of the localization of a protein in  
30 a cell, e.g., tagging elements such as fluorescent protein coding sequences, such as the GFP coding sequences; (c) elements necessary for constitutive, bicistronic

expression in mammalian cells, e.g., IRES sites, in combination with selectable markers, e.g. antibiotic resistance, fluorescent protein, etc. ; (d) elements necessary for inducible expression of the gene of interest on an expression vector, e.g. inducible promoters such as the tet-responsive promoter, etc.; (e)

5 elements that provide for retroviral expression vectors; and the like.

In addition to the above requisite and optional elements, the acceptor vectors further include at least one splice site. Two representative but non-limiting embodiments are now described further. In a first embodiment, the acceptor vector includes a splice acceptor site positioned 3' of the single sequence specific  
10 recombinase site of the vector. More precisely, this splice acceptor sequence is placed 3' of the acceptor partial selectable marker sequence. This embodiment finds use in applications where one wishes to produce expression vectors in which the gene of interest is not expressed as a fusion with 3' sequence specific recombinase site encoded domains, etc. In a second representative  
15 embodiment, the acceptor vector further includes a splice donor site which is positioned 5' of the single sequence specific recombinase site, where this embodiment finds use in those situations where one wishes to produce an expression vector in which the gene of interest is expressed as a protein that does not include either N or C-terminal residues encoded by the 5' and 3'  
20 sequence specific recombinase sites.

#### *Product Vector Generation with a Recombinase*

As mentioned above, in the subject methods the donor and acceptor  
25 vectors are contacted with a recombinase under conditions sufficient for site specific recombination to occur, specifically under conditions sufficient for a recombinase mediated recombination event to occur that produces the desired intron containing product vector, where product vector production is accomplished without cutting or ligation of the donor and acceptor vectors with restriction  
30 endonucleases and nucleic acid ligases. The contact may occur under in vitro or in vivo conditions, as is desired and/or convenient.

In many embodiments, an aqueous reaction mixture is produced by combining the donor and acceptor vectors and the recombinase with water and other requisite and/or desired components to produce a reaction mixture that, under appropriate conditions, results in production of the desired expression vector. The various components may be combined separately or simultaneously, depending on the nature of the particular component and how the components are combined. Conveniently, the components of the reaction mixture are combined in a suitable container. The amount of donor and acceptor vectors that are present in the reaction mixture are sufficient to provide for the desired production of the expression vector product, where the amounts of donor and acceptor vector may be the same or different, but are in many embodiments substantially the same if not the same. In many embodiments, the amount of donor and acceptor vector that is present in the reaction mixture ranges from about 50 ng to 2 µg, usually from about 100 ng to 500 ng and more usually from about 150 ng to 300 ng, for a reaction volume ranging from about 5 µl to 1000 µl, usually from about 10 µl to 50 µl.

The recombinase that is present in the reaction mixture is one that provides for recombination of the donor and acceptor vectors, i.e. one that recognizes the recombinase recognition sites on the donor and acceptor vectors.

As such, the recombinase employed will vary, where representative recombinases include, but are not limited to: recombinases, transposes and integrases, where specific recombinases of interest include, but are not limited to: Cre recombinase (the cre gene has been cloned and expressed in a variety of hosts, and the enzyme can be purified to homogeneity using standard techniques known in the art-- purified Cre protein is available commercially from CLONTECH, Novagen, NEB, and others); FLP recombinase of *S. cerevisiae* that recognizes the frt site; Int recombinase of bacteriophage Lambda that recognizes the att site; xerC and xerD recombinases of *E.coli*, which together form a recombinase that recognizes the dif site. the Int protein from the Tn916 transposon; the Tn3 resolvase, the Hin recombinase; the Cin recombinase; the immunoglobulin recombinases; and the like. While the amount of recombinase present in the

reaction mixture may vary depending on the particular recombinase employed, in many embodiments the amount ranges from about 0.1 units to 1250 units, usually from about 1 unit to 10 units and more usually from about 1 unit to 2 units, for the above described reaction volumes. The aqueous reaction mixture may include

5 additional components, e.g., a reaction buffer or components thereof, e.g., buffering compounds, such as Tris-HCl; MES; sodium phosphate buffer, sodium acetate buffer; and the like, which are often present in amounts ranging from about 10 mM to 100 mM, usually from about 20 mM to 50 mM; monovalent ions, e.g., sodium, chloride, and the like, which are typically present in amounts

10 ranging from about 10 mM to 500 mM, usually from about 30 mM to 150 mM; divalent cations, e.g., magnesium, calcium and the like, which are often present in amounts ranging from about 1 mM to 20 mM, usually from about 5 mM to 10 mM; and other components, e.g., BSA, EDTA, spermidine and the like; etc (where the above amount ranges are provided for the representative reaction volumes

15 described above). As the reaction mixtures are aqueous reaction mixtures, they also include water.

The subject reaction mixtures are typically prepared at temperatures ranging from about 0-4°C, e.g., on ice, to minimize enzyme activity. Following reaction mixture preparation, the temperature of the reaction mixture is typically raised to a temperature that provides for optimum or maximal recombinase activity, and concomitantly expression vector production. Often, in this portion of the method the temperature will be raised to a temperature ranging from about 4 °C to 37 °C, usually from about 10 °C to 25 °C , where the mixture will be maintained at this temperature for a period of time sufficient for the desired

20 amount of expression vector production to occur, e.g., for a period of time ranging from about 5 mins to 60 mins, usually from about 10 mins to 15 mins. Following the incubation period, the reaction mixture is subjected to conditions sufficient to inactivate the recombinase, e.g., the temperature of the reaction mixture may be raised to a value ranging from about 65 °C to 70 °C for a period of time ranging

25 from about 5 mins to 10 mins.

Alternatively, contact of the donor and acceptor vectors with the recombinase may occur *in vivo*, where the donor and acceptor vectors are introduced in a suitable host cell that expresses a recombinase. In this embodiment, the recombination between the donor and acceptor vectors may be

5       accomplished *in vivo* using a host cell that transiently or constitutively expresses the appropriate site-specific recombinase (e.g., Cre recombinase expressed in the bacterial strain BNN132, available from CLONTECH). pDonor and pAcceptor, i.e., the donor and acceptor vectors respectively, are co-transformed into the host cell using a variety of methods known in the art (e.g., transformation of cells made

10      competent by treatment with CaCl<sub>2</sub>, electroporation, etc.). The co-transformed host cells are grown under conditions which select for the presence of the recombinant-functional selectable marker created by recombination of pDonor with the pAcceptor (e.g., growth in the presence of chloramphenicol and sucrose when the pDonor vector contains the SacB negative selection marker on the non

15      donor fragment and all or part of the chloramphenicol resistance gene open reading frame and pAcceptor may also contain a promoter necessary for expression of the chloramphenicol open frame). Plasmid DNA is isolated from host cells which grow in the presence of the selective pressure and is subjected to restriction enzyme digestion to confirm that the desired recombination event

20      has occurred.

The present invention also provides a method for the *in vitro* recombination of nucleic acid constructs, comprising the steps of: a) providing i) a donor nucleic acid construct comprising a donor-partial selectable marker, two donor sequence-specific recombinase target sites each having a defined 5' to 3' orientation and

25      wherein the donor sequence-specific recombinase target sites are placed in the donor construct such that they have the same 5' to 3' orientation, and a unique restriction enzyme site or polylinker, the restriction enzyme site or polylinker being located 3' of the first-donor sequence-specific recombinase target site and 5' of the second-donor sequence-specific recombinase target site; (ii) an acceptor

30      nucleic acid construct comprising an origin of replication, an acceptor sequence-specific recombinase target site having a defined 5' to 3' orientation, a first

promoter adjacent to the 5' end of the acceptor sequence-specific recombinase target site, and an acceptor-partial selectable marker, wherein the acceptor-partial selectable marker is capable of recombining with the donor-partial selectable marker from the donor construct to create a recombinant-functional selectable marker in a final desired recombination construct; b) contacting the donor and acceptor constructs *in vitro* with a site-specific recombinase under conditions such that the desired donor fragment recombines with the acceptor construct to form a final desired recombination construct.

The present invention further provides a method for the recombination of nucleic acid constructs in a host, comprising the steps of: a) providing i) a donor nucleic acid construct comprising a donor-partial selectable marker, two donor sequence-specific recombinase target sites each having a defined 5' to 3' orientation and wherein the donor sequence-specific recombinase target sites are placed in the donor construct such that they have the same 5' to 3' orientation, and a unique restriction enzyme site or polylinker, the restriction enzyme site or polylinker located 3' of the first-donor sequence-specific recombinase target site and 5' of the second-donor sequence-specific recombinase target site; (ii) an acceptor nucleic acid construct comprising an origin of replication, an acceptor sequence-specific recombinase target site having a defined 5' to 3' orientation, a first promoter adjacent to the 5' end of the acceptor sequence-specific recombinase target site, and an acceptor-partial selectable marker, wherein the acceptor-partial selectable marker is capable of recombining with the donor-partial selectable marker from the donor to create a recombinant-functional selectable marker in a final desired recombination construct; and iii) a host cell expressing a site-specific recombinase; b) introducing the donor and acceptor constructs into the host cell under conditions such that the desired donor fragment recombines with the acceptor construct to form the final desired recombination construct which is capable of imparting the ability to the host cell to grow in selective growth medium.

The above methods of producing expression vectors can be employed to rapidly produce a plurality of different expression vectors that are distinct from

each other but carry the same coding sequence of interest from a single, original type of donor vector. In other words, the subject methods can be used to rapidly clone a nucleic acid of interest from an initial vector into a plurality of expression vectors. By plurality is meant at least 2, usually at least 5, and more usually at 5 least 10, where the number may be as high as 20, 96 or more. The methods can be performed by one person in a period of time that is a fraction of what it would take by that person of skill in the art to produce the same number and variety of expression vectors using traditional cutting and ligation protocols, where the increase in efficiency obtained by the subject methods is at least about 6 fold, 10 usually at least about 15 fold and more usually at least about 30 fold.

### *The Resultant Product Vector*

The above steps result in the production of an intron containing product vector (i.e. a vector that includes one or more, e.g., one or two, spliceable introns) from donor and acceptor vectors, and in certain embodiments from a portion of one of these vectors and the entirety of the other of these vectors, e.g., from a portion of the donor vector and the entirety of the acceptor vector, where by portion is meant the part of the donor vector that lies 3' of the first donor sequence-specific recombinase site and 5' of the second donor sequence-specific recombinase site. The size of the product vector may vary, depending on the nature of the vector. Where the vector is a plasmid, the size of the expression vector may range from about 3 kb to 20 kb, usually from about 4 kb to 8 kb.

The resultant product vector in many embodiments is characterized in that 25 it includes two recombinase recognition sites, i.e., a first and second recombinase recognition site, oriented in the same direction. The distance between the first and second recombinase sites, specifically the distance between the 3' end of the first recombinase site and the 5' end of the second recombinase site, ranges in many embodiments from about 100 bp to 100 kb, usually from about 500 bp to 20 30 kb, depending on whether the coding sequence of a protein of interest or just a restriction site/multiple cloning site, is present between the first and second

recombinase recognition sites. The portion of the vector that lies in this inter recombinase region, i.e. 3' of the first recombinase site and 5' of the second recombinase site, typically makes up from about 2 % to 85%, usually from about 20% to 60 % of the entire expression vector.

5        In many embodiments, the expression vector is further characterized in that 5' of the first recombinase site is a first promoter, 3' of the first recombinase site is at least one restriction site; and the second recombinase site located inside a functional selectable marker, i.e., it is flanked by disparate portions or sub-parts of a selectable marker expression module or cassette (e.g., a promoter and a

10      coding sequence), where the second recombinase site is present between the two sub-parts of the selectable marker in a manner such that the selectable marker is functional, i.e., the coding sequence of the selectable marker is expressed. In other words the expression vector includes a selectable marker expression cassette or module made up of a promoter and coding sequence that

15      flank the second recombinase site. In many embodiments, the second recombinase site is flanked by a promoter on its 3' end and a coding sequence of the selectable marker on its 5' end. In this embodiment, the first and second promoters, located 5' of the first recombinase site and 3' of the second recombinase site, respectively, are oriented in opposite directions.

20      The expression vector is further characterized by having at least one restriction site, and generally a multiple cloning site, located between the first and second recombinase sites. In many embodiments, located between the first and second recombinase sites, and flanked by two restriction sites, which may or may not be the same, is a nucleic acid of interest, i.e., gene of interest, that includes a

25      coding sequence for a protein of interest whose expression from the expression vector is desired. In these embodiments, the first promoter 5' of the first recombinase site and the coding sequence for the protein of interest are arranged on either side of the first recombinase site such that they form an expression module or cassette that expresses the encoded protein, i.e., the coding sequence

30      and first promoter flank the first recombinase site in manner such that they are operably linked.

In addition to the above features, the expression vector further includes at least one origin of replication that provides for replication in the host or hosts into which it is placed or transformed during use. Origins of replication of interest include, but are not limited to, those described above in connection with the donor and acceptor vectors.

In certain embodiments, the product vector contains a gene or DNA sequence of interest inserted into the unique restriction enzyme site or polylinker such that the gene or DNA sequence of interest is under the control of the first promoter. The gene or DNA sequence of interest is joined to the 3' end of the first-recombinant sequence-specific recombinase target site such that a functional transcriptional unit is formed so that the gene or DNA sequence of interest is expressed as a protein driven by the first promoter of the acceptor construct. In a more preferred embodiment, the gene of interest is joined to the 3' end of the first-recombinant sequence-specific recombinase target site such that a functional translational reading frame is created wherein the gene or DNA sequence of interest is expressed as a fusion protein with an affinity domain or tag sequence derived from the acceptor plasmid and under the expression control of the first promoter of the acceptor construct.

In another preferred embodiment, the gene of interest is joined to the donor splice site such that when the intron is spliced out of the resultant mRNA, the gene of interest is fused in frame to a C-terminal tag derived from the acceptor vector.

In certain embodiments, the product vector further comprises an acceptor-functional selectable marker gene derived from the acceptor construct. If an acceptor-functional selectable marker gene is present in addition to the newly-created recombinant-functional selectable marker, the acceptor-functional selectable marker is a different selectable marker from the newly-created recombinant-functional selectable marker. The present invention should not be limited by the nature of the selectable marker genes chosen; the marker genes may result in positive or negative selection and may be chosen from the group including, but not limited to, the chloramphenicol resistance gene, the ampicillin

resistance gene, the tetracycline resistance gene, the kanamycin resistance gene, the streptomycin resistance gene, the strA gene and the sacB gene.

In addition to the above features, the product vector further includes at least one, and typically one to two, spliceable introns. The one or more introns  
5 may be positioned anywhere in the product vector. In certain representative embodiments, the 3' recombinase recognized site is present in an intron. In other representative embodiments, the 5' recombinase recognized site is present in an intron. In yet other representative embodiments, both the 5' and 3' recombinase recognized sites are present in introns.

10

## UTILITY

The subject methods find use in a variety of different applications, where  
15 such applications are generally those protocols and methods in which the transfer of a nucleic acid of interest from one vector to another, e.g., the cloning of a nucleic acid from an initial vector into a final vector, is desired. As such, the subject methods are particularly suited for use in cloning nucleic acids of interest, including whole libraries, from an initial vector into an expression vector, where  
20 the product vector may be functionalized to express the polypeptide or protein encoded by the nucleic acid of interest located on it in a variety of different desired environments and/or under desired conditions, e.g., in a cell of interest, in response to a particular stimulus, tagged by a detectable marker, etc.

As such, the product vectors produced by the subject methods find use in  
25 a variety of different applications, including the study of polypeptide and protein function and behavior, i.e., in the characterization of a polypeptide or protein, either known or unknown; and the like. In the broadest sense, the subject methods find application in any method where traditional digestion and ligation protocols are employed to transfer or clone a nucleic acid from one vector to  
30 another, e.g., cloning digestion and ligation protocols, where the expression vectors produced by the subject methods find use in research applications, as

well as other applications, e.g., protein production applications, therapeutic applications, and the like.

Depending on the location of the one or more introns in the product vectors, the product vectors find use in the expression of non-fusion proteins,

5 e.g., proteins free of residues at their N- and C-termini that are encoded by sequence specific recombinase sites; N-and or C-termini tagged proteins, etc.

## SYSTEMS

10 Also provided are systems for use in practicing the subject methods. The subject systems at least include a donor vector and an acceptor vector as described above. In addition, the subject systems may include a recombinase which recognizes the recombinase sites present on the donor and acceptor vectors. The systems may also include, where desired, a host cell, e.g., in vivo methods of expression vector production, as described above. Other components of the subject systems include, but are not limited to: reaction buffer, controls, etc.

## LIBRARIES

20 Also provided are nucleic acid libraries cloned into donor and/or acceptor vectors of the subject invention. These nucleic acid libraries are made up of a plurality of individual donor/acceptor vectors where each distinct constituent member of the library has a different nucleic acid portion or component, e.g., genomic fragment, cDNA, of an original whole nucleic acid library, i.e., fragmented genome, cDNA collection generated from the total or partial mRNA of an mRNA sample, etc. In other words, the libraries of the subject invention are nucleic acid libraries cloned into donor or acceptor vectors according to the subject invention, where the nucleic acid libraries include, but are not limited to, 25 genomic libraries, cDNA libraries, etc. Specific donor/acceptor libraries of interest include, but are not limited to: Human Brain Poly A+ RNA; Human Heart Poly A+ 30

RNA; Human Kidney Poly A+ RNA; Human Liver Poly A+ RNA; Human Lung Poly A+ RNA; Human Pancreas Poly A+ RNA; Human Placenta Poly A+ RNA; Human Skeletal Muscle Poly A+ RNA; Human Testis Poly A+ RNA; Human Prostate Poly A+ RNA and the like. With donor libraries according to the subject invention, the  
5 subject methods permit the rapid exchange of either individual clones of interest, groups of clones or potentially an entire cDNA library to a variety of expression vectors.

## KITS

10

Also provided are kits for use in practicing the subject methods. The subject kits at least include at least one donor vector and a recombinase that recognizes the recombinase sites of the donor vector. The subject kits may further include other components that find use in the subject methods, e.g.,  
15 acceptor vectors; reaction buffers, positive controls, negative controls, etc.

20

In addition to the above components, the subject kits will further include instructions for practicing the subject methods. These instructions may be present in the subject kits in a variety of forms, one or more of which may be present in the kit. One form in which these instructions may be present is as printed information on a suitable medium or substrate, e.g., a piece or pieces of paper on which the information is printed, in the packaging of the kit, in a package insert, etc. Yet another means would be a computer readable medium, e.g., diskette, CD, etc., on which the information has been recorded. Yet another means that may be present is a website address which may be used via the internet to access the information at a removed site. Any convenient means may be present  
25 in the kits.

30

The following examples are offered by way of illustration and not by way of limitation.

B, F & F Ref: CLON-069  
Clontech Ref: P-90

F:\DOCUMENT\CLON\069\patent application.doc

## EXPERIMENTAL

### Example 1. Representative Protocols

5

#### A.

Figure 5 provides a flow diagram of a representative recombinase based method according to the subject invention.

10 B.

In order to test the utility of intron-splicing to enable tagging of a protein of interest in a donor vector with a peptide tag or protein in an acceptor vector, a Donor and Acceptor vector capable of splicing were built using standard molecular biology techniques. The Donor vector was called pDNR-Dual. A map of this vector is provided in Figure 1 and its sequence is provided below as SEQ ID NO:01. The Acceptor vector was called pLPS-EGFP. A map of this vector is provided in Figure 2 and its sequence is provided below as SEQ ID NO:02. Further, a luciferase test gene was cloned, using standard techniques into the MCS of pDNR-Dual at the Sall and Apa I sites, so as to generate pDNR-Dual-Luc. A map of this vector is provided in Figure 3 and the sequence of this vector is provided below as SEQ ID NO:03. In so doing, the Luciferase gene was placed such that it had no stop codon and such that it would be in-frame with the EGFP tag present in pLPS-EGFP following Cre/Lox-based transfer from the Donor to the Acceptor.

25

The pDNR-Dual-Luc and pLPS-EGFP vectors were then recombined in vitro using Cre according to methods described in Clontech's Creator User Manual (Clontech Laboratories Inc., Palo Alto CA) (see also the methods disclosed in U.S. Application Serial No. 09/616,651, the disclosure of which is herein incorporated by reference), and an aliquot of the reaction was transformed 30 into competent E. coli. Following selection on chloramphenicol and sucrose plates, recombinant clones were isolated and confirmed by standard restriction

mapping and sequencing to encode the expected recombinant molecule, having the luciferase gene from the donor vector transferred to the acceptor vector. This vector is called pLPS-Luc-EGFP. A map of this vector is provided in Figure 4 and the sequence of this vector is provided below as SEQ ID NO:04. This construct  
5 thus has both a splice donor sequence, provided from the donor vector, and a splice acceptor sequence, provided by the acceptor vector. Together, these create an artificial intron between the 3' end of the luciferase gene and the 5' end of the EGFP Tag. This intron being composed of the chloramphenicol open reading frame, the second LoxP site, and the ampicillin promoter sequence.

10 To test if this construct would generate a properly spliced mRNA, so enabling expression of a luciferase EGFP fusion protein, the pLPS-Luc-EGFP vector was then transfected into HEK293 cells using standard procedures known to the art. For comparison, the HEK293 cells were also transfected with a pLuc-EGFP construct. This construct was made by cloning the luciferase gene (without stop codon) in-frame with EGFP into the pEGFP-N1 vector (available from Clontech Laboratories, Inc. Palo Alto CA) using standard molecular biology techniques.  
15

Twenty-four hours after transfection, the cells were examined for EGFP fluorescence using a fluorescence microscope. Both the splicing construct  
20 (pLPS-Luc-EGFP) and the direct luciferase-EGFP fusion (pLuc-EGFP) showed equivalent EGFP expression over untransfected control cells.

Extracts of the cells were then made and analyzed by western blotting using an anti-luciferase antibody. Again, both the splicing construct (pLPS-Luc-EGFP) and the direct luciferase-EGFP fusion (pLuc-EGFP) showed equivalent  
25 expression of the luciferase-EGFP fusion protein. A further analysis of total RNA extracted from cells transfected with the splicing construct (pLPS-Luc-EGFP) by Northern blotting, demonstrated that the mRNA generated from the construct was being efficiently spliced to remove the chloramphenicol sequences.

30

### Example 2. Vector Sequence Information

## A. pDNR-dual

2941 gaaaatgccg atatccattt ggcattgacg tcaggtggca ctttcgggg aatgtgcgc  
 3001 ggaaccctta ttttttattt tttctaaata cattcaaata tgtatccgct catgagacaa  
 3061 taaccctgat aaatgctca ataatattga aaaaggaga gtatgagtat tcaacatttc  
 3121 cgtgtcgccc ttattccctt ttttgcggca ttttgccttc ctgttttgc tcacccagaa  
 3181 acgctggta aagtaaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa  
 3241 ctggatctca acagcggtaa gatcctttag agtttcggcc cccaagaacg ttttccaatg  
 3301 atgagcactt ttaaaggttct gctatgtggc gcggtattat cccgtattga cgccgggcaa  
 3361 gagcaactcg gtcgcccatt acactattct cagaatgact tgggttagta ctcaccagtc  
 3421 acagaaaagc atcttacgga tgcatgaca gtaagagaat tatgcagtgc tgccataacc  
 3481 atgagtgata acactgcggc caacttactt ctgacaacga tcggaggacc gaaggagcta  
 3541 accgctttt tgcacaacat gggggatcat gtaactcgcc ttgategttg ggaaccggag  
 3601 ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtgc aatggcaaca  
 3661 acgttgcgc aactattaac tggcgaacta cttactctag cttccggca acaattaata  
 3721 gactggatgg aggccgataa agttgcagga ccacttctgc gctcgccct tccggctggc  
 3781 tggtttattt ctgataaaatc tgagccgg gaggcgtggg ctcgcgttat cattgcagca  
 3841 ctggggccag atggtaagcc ctcccgatc gtatgtatct acacgacggg gagtcaggca  
 3901 actatggatg aacgaaatag acagatcgct gagatagggtg cctcactgtat taagcattgg  
 3961 taactgtcag accaagttt ctcataatata ctttagattt attaaaact tcatttttaa  
 4021 tttaaaagga tctaggtgaa gatcctttt gataatctca tgacaaaaat cccttaacgt  
 4081 gagtttcgt tccactgagc gtcagacccc gtagaaaaaga tcaaaggatc ttcttgagat  
 4141 ccttttttc tgcgcgtaat ctgctgctt gaaacaaaaa aaccaccgt accagcgggt  
 4201 gttgtttgc cggatcaaga gctaccactt cttttccga aggttaactgg cttcagcaga  
 4261 ggcagataac caaatactgt tcttctagtg tagccgtagt taggccacca cttcaagaac  
 4321 tctgttagcac cgcctacata cctcgctctg ctaatcctgt taccagtggc tgctgccagt  
 4381 ggcgataagt cgtgttttac cgggttggac tcaagacgt agttaccggta taaggcgcag  
 4441 cggcgggctt gAACGGGGGG ttctgtgcaca cagccccagct tggagcgaac gacccatcacc  
 4501 gaactgagat acctacagcg ttagctatga gaaagcgcac cgcttcccgaa agggagaaag  
 4561 gcccggatgt atccggtaag cggcagggtc ggaacaggag agccacggag ggagcttcca  
 4621 gggggaaacg cctggtatct ttatagtcc tgcgggtttc gccacctctg acttgacgt  
 4681 cgatTTTGT gatgctgtc agggggccgg agcctatgaa aaaaacgcac caacgcggcc  
 4741 ttttacggt tcctggccct ttgctggcct tttgctcaca ttttctttcc tgcttatcc  
 4801 cctgattctg tggataaccg tattaccgc ttacgcgtgt aaaaacgcacgg ccagtagatc  
 4861 tgtaatacga ctcactatag ggcgctagct gctcgccgc gccgaacgcg cggcgcac  
 4921 gagtcagtga gcggagaa (SEQ ID NO:01)

## B. pLPS-EGFP

1 tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg  
 61 cgttacataa cttacgtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt  
 121 gacgtcaata atgacgtatg ttcccatagt aacgcoaata gggactttcc attgacgtca  
 181 atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc  
 241 aagtaacccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta  
 301 catgaccta tgggacttcc ctacttggca gtacatctac gtattagtca tcgcttattac  
 361 catggtgatg cgggtttggc agtacatcaa tggcgtgga tagcggtttgc actcacgggg  
 421 atttccaagt ctccacccca ttgacgtcaa tggagtttgc ttttggcacc aaaaatcaacg  
 481 ggactttcca aatgtcgta acaactccgc cccattgacg caaatggcg gtaggcgtgt  
 541 acggtgggag gtctatataa gcagagctgg ttttagtgaac cgtcagatcc gctagataa  
 601 cttcgatag catacattat acgaagttt agatccaata ttattgaacg atttattcagg  
 661 gttattgtct catgagcgga tacatattt aatgtattt gaaaaataaaa caaatagggg  
 721 ttccgcgcac atttccccca aaagtgcac ctgacgtgga tctcgagctc aagcttcgaa  
 781 ttcaagggtt cttgacaat atcataactta tccgtccct ttttttccca cagctaccgg  
 841 tcgcgagcaa gggcgaggag ctgttcaccc ggggtgggcc catcctggtc gagctggacg  
 901 ggcacgtaaa cggccacaag ttacgcgtgt cggcggagg gggggcgat gcccaccc  
 961 gcaagctgac cctgaagttc attcgcacca cggcgaagct gcccgtgccc tggccaccc  
 1021 tcgtgaccac cctgaccatc ggcgtgcagt gttcagccg ctaccccgac cacatgaagg

5	1081	agcacgactt	cttcaagtcc	gccatccccg	aaggctacgt	ccaggagcgc	accatctct
	1141	tcaaggacga	cggcaactac	aagaccgcg	ccgaggtgaa	tttcgagggc	gacaccctgg
	1201	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cgcaacatc	ctggggcaca
	1261	agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaaag	cagaagaacg
	1321	gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cgccagcgtg	cagctccgg
	1381	accactacca	gcagaacacc	cccatcgccg	acggcccg	gctgtgccc	gacaaccact
	1441	acctgagcac	ccagtccgcc	ctgagcaaag	accccaacga	gaagcgcgt	cacatggcc
10	1501	tgctggagtt	cgtgaccgccc	gccggatca	ctctcgcat	ggacgagctg	tacaagtaaa
	1561	gccccccgca	ctctagatca	taatcagcca	taccacattt	gtagaggttt	tacttgcttt
	1621	aaaaaacctc	ccacacctcc	ccctgaacct	gaaacataaa	atgaatgcaa	tttgttgtgt
	1681	taacttgttt	attgcagctt	ataatggta	caaataaaagc	aatagcatca	caaatttcac
	1741	aaataaaagca	ttttttcac	tgcattctag	tttgtggttt	tccaaactca	tcaatgtatc
	1801	ttaaggcgt	aattgtaaagc	gttaatattt	tgttaaaattt	cgcgttaaat	ttttgttaaa
15	1861	tcagctcatt	tttaaacc	taggccc	tcggcaaaat	cccttataaa	tcaaaaagaat
	1921	agaccgagat	agggttgagt	gttgtccag	tttggAACAA	gagtccacta	ttaaagaacg
	1981	tggactccaa	cgtcaaaagg	cgaaaaaaccg	tctatcaggg	cgatggccca	ctacgtgaac
	2041	catcacccta	atcaagttt	ttggggtcga	ggtgcgtaa	agcactaaat	cggaaacccta
	2101	aagggagccc	ccgatttaga	gcttgacggg	gaaagccggc	gaacgtggcg	agaaaggaaag
20	2161	ggaagaaagc	gaaaggagcg	ggcgttaggg	cgctggcaag	tgttagcgtc	acgctgcgcg
	2221	taaccaccac	acccgcgcg	cttaatgcgc	cgctacaggg	cgcgtcaggt	ggcactttc
	2281	ggggaaatgt	gcmcggaaacc	cctatttgtt	tatTTTCTA	aatacattca	aatatgtatc
	2341	cgctcatgag	acaataaccc	tgataaaatgc	ttcaataata	ttgaaaaaagg	aagagtctg
	2401	aggcggaaag	aaccagctgt	ggaatgtgt	tcagttaggg	tgtgaaaagt	ccccaggc
	2461	cccagcaggc	agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	ggtgtggaaa
	2521	gtccccaggc	tccccagcag	gcagaagtat	gcaaagcatg	catctcaatt	agtcagcaac
	2581	catagcccc	ccccctaactc	cgccccatccc	gccccctaact	ccgcccagtt	ccgcccattc
	2641	tccgcccat	ggctgactaa	ttttttttat	ttatgcagag	gccgaggccc	cctcggcctc
	2701	tgagctattc	cagaagtagt	gaggaggctt	ttttggaggc	ctagctttt	gcaaagatcg
30	2761	atcaagagac	aggatgagga	tcgtttcgca	tgattgaaca	agatggattt	cacgcagtt
	2821	ctccggccgc	ttgggtggag	aggctattcg	gctatgactg	ggcacaacag	acaatcgct
	2881	gctctgatgc	cgccgtgttc	cggtgtcag	cgccaggccg	cccgttctt	tttgtcaaga
	2941	ccgaccctgtc	cggtgccttg	aatgaactgc	aagacgaggc	agcgcggcta	tcgtggctgg
	3001	ccacgacggg	cgttccttgc	gcagctgtgc	tcgacgttgt	cactgaagcg	ggaaggact
35	3061	ggctgctatt	gggcgaagt	ccggggcagg	atctcctgtc	atctcacctt	gctcctccgg
	3121	agaaagtatc	catcatggct	gatgcaatgc	ggcggctgca	tacgcttgc	ccggctaccc
	3181	gcccattcga	ccaccaagcg	aaacatcgca	tcgagcgcagc	acgtactcgg	atgaaaggccg
	3241	gtcttgtcga	tcaggatgt	ctggacgaag	agcatcaggg	gctcgcgcca	ccgaaactgt
	3301	tcgcccaggct	caaggcgcagc	atgcccgcac	gcgaggatct	cgtcgtgacc	catggcgatg
	3361	cctgcttgcc	gaatatcatg	gtggaaaatg	gccgcctttc	tggattcatc	gactgtggcc
40	3421	ggctgggtgt	ggcggaccgc	tatcaggaca	tagcgttggc	taccgtgtat	attgctgaag
	3481	agcttggcg	cgaatgggc	gaccgttcc	tcgtgcttta	cggtatcgcc	gctcccgatt
	3541	cgcagcgcat	cgccttctat	cgccttctg	acgagttctt	ctgagcggga	ctctggggtt
	3601	cgaaatgacc	gaccaagcga	cgcccaacct	gccatcacga	gattcgatt	ccaccgcgc
	3661	cttctatgaa	aggtgggct	tcggaatcgt	tttccggac	gccgctgga	tgatcctcca
45	3721	gccccgggat	ctcatgctgg	agttttcgc	ccaccctagg	gggaggctaa	ctgaaacacg
	3781	gaaggagaca	ataccggaa	gaaccgcgc	tatgcggca	ataaaaaagac	agaataaaaac
	3841	gcacgggtt	gggtcgttt	ttcataaaacg	cggggttcgg	tcccagggt	ggcactctgt
	3901	cgataccca	ccgagacccc	attggggcca	atacgcccc	gtttcttct	tttccccacc
	3961	ccacccccc	agttcggtt	aaggcccagg	gctcgcagcc	aacgtcgcccc	ccgcaggccc
50	4021	tgccatagcc	tcaggttact	catatatact	ttagattgt	ttaaaaacttc	attttaatt
	4081	taaaaggatc	taggtgaaga	tccttttga	taatctcatg	acccaaatcc	cttaacgtga
	4141	gttttcgttc	cactgagcgt	cagacccgt	agaaaaagatc	aaagatctt	cttgagatcc
	4201	ttttttctg	cgcgtatct	gctgttgc	aacaaaaaaaa	ccaccgc	cagcgttggt
	4261	ttgtttccgc	gatcaagagc	taccaactct	ttttccgaa	gtactggct	tcagcagagc
	4321	gcagatacca	aatactgtcc	ttctagtgt	gccgtagtt	ggccaccact	tcaagaactc
55	4381	tgttagcaccg	cctacataacc	tcgctctgt	aatcctgtt	ccagtggt	ctgcccq

4441 cgataagtgc tgtcttaccg gggtggactc aagacgatag ttacccggata aggccgcagcg  
4501 gtcgggctga acggggggtt cgtgcacaca gcccagctg gagcgaacga cctacaccga  
4561 actgagatac ctacacgcgtg agctatgaga aagcgccacg ctccccgaag ggagaaaaggc  
4621 ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg agttccagg  
4681 gggaaacgcc tggtatcttt atagtcctgt cgggttgcgc cacctctgac ttgagcgtcg  
4741 atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgcccagca acgcggccctt  
4801 tttacggttc ctggccttt gctggccttt tgctcacatg ttcttcctg cgttatcccc  
4861 tqattctgtq qataaccqta ttaccqccat qcat (SEQ ID NO:02)

### 10 C. pDNR-Dual-Luc

1 gcggccgcat aacttcgtat agcatacatt atacgaagtt atcagtgcac accatggaa  
61 acgccaaaaa cataaaagaaa ggcccggcgc cattctatcc tctagaggat ggaaccgctg  
121 gagagcaact gcataaggct atgaagagat acgcccgtt tcctgaaca attgctttt  
181 cagatgcaca tatcgaggtg aacatcacgt acgcggaaata cttegaaatg tccggttcgg  
241 tggcagaagc tatgaaacga tatgggctga atacaatca cagaatcgct gtatgcagtg  
301 aaaactctct tcaattctt atgcccgtt tggcgcgtt atttatcgga gttcagtt  
361 cgccccgaa cgacatttat aatgaacgtg aattgctcaa cagtagaac atttcgcagc  
421 ctaccgtat gttgtttcc aaaaagggt tgcaaaaaat ttgaacgtg caaaaaaaaaat  
481 taccataat tcagaaaattt attatcatgg attctaaaac ggattaccag ggatttcagt  
541 cgatgtacac gtcgtcaca tctcatctac ctccccgtt taatgagtc gatgttgc  
601 cagagtcctt tgatcgtaaaaacaattt cactgataat gaattcctt gatctactg  
661 ggttaccaa ggggtggcc ctccgcata gaactgcctg cgtcagattc tcgcatgc  
721 gagatcctat tttggcaat caaatcattt cggatactgc gatTTAAGT gtttccat  
781 tccatcacgg ttttgaatg tttactacac tcggatattt gatatgtgga tttcgagtc  
841 tcttaatgtat tagatttga aagagactgt tttacgatc cttcaggat tacaaaattt  
901 aaagtgcgtt gctagtagcca accctatTTT cattctcgc caaaagcact ctgattgaca  
961 aatacgtattt atctaatttta cacgaaattt cttctggggg cgcacctt tcgaaagaag  
1021 tcgggaaagc ggttgcaaaaa cgcttccatc ttccaggat acgacaagga tatggc  
1081 ctgagactac atcagctatt ctgattacac ccggggggg tgataaaccg ggcgcggc  
1141 gtAAAAGTTGT tccatTTTTT gaagcgaagg ttgtggatct ggataccggg aaaacgc  
1201 gcgttaatca gagaggcgaa ttatgtgtca gaggacctat gattatgtcc gttatgtaa  
1261 acaatccgga agcgaccaac gccttgattt acaaggatgg atggctacat tctggagaca  
1321 tagcttactg ggacgaagac gaacacttct tcatagttga ccgcttgaag tcttaatta  
1381 aatacaaagg atatcagggtg gccccgcgtg aatttggatc gatattgtt caacaccccc  
1441 acatctcga cgcggcggtg gcaggcttc ccgacatgc cgcgtgaa cttccggcc  
1501 ccgttgttgt ttggagcac ggaaagacga tgacggaaaa agagatcgat gattacgtc  
1561 ccagtcaagt aacaaccgcg aaaaagtgc gggaggatgt gttgttgc gacgaagtac  
1621 cgaaagggtct taccggaaaaa ctcgacgc aaaaaatcag agatcctc ataaaggcca  
1681 agaaggcgg aaagtccaaa ttgaggatcc gggcccaggat ggtgttcat aatcataatc  
1741 ataatcataa tcataatcac aactagccata ggagatccctg gtcatgacta gtgttggat  
1801 tctcaccaat aaaaaacgcgc cggcggcaac cgagcgttct gaacaaatcc agatggagtt  
1861 ctgaggtcat tactggatct atcaacagga gtcacgc gtcgatatac aaattacgc  
1921 ccgcctgccc actcatcgca gtactgttgc aattcattaa gcattctgccc gacatggaa  
1981 ccatcacaaa cggcatgtat aacctgaatc gccagcggca tcagcacctt gtcgc  
2041 gtataatatt tgcccatggt gaaaacgggg gcaagaagt tgcatttgcattt ggcacgtt  
2101 aatcaaacc tggtgaaact cacccaggaa ttggctgaga cgaaaaacat attctcaata  
2161 aaccctttag gggaaataggc cagggtttca cctgtacaccc ccacatctt cgaatataatg  
2221 tggtagaaact gccggaaatc gtcgtgttgc tcaactccaga gcgatggaaa ctttc  
2281 tgctcatggaa aacgggtgt acaagggtga acactatccc atatcaccag ctcaccgt  
2341 ttcattgcatacgaaattt cggatgagca ttcatcaggc gggcaagaat gtgaaataaag  
2401 gcccggataaa acttgcgtt attttctt acggcttta aaaaaggccgt aatatccagc  
2461 tgaacgtct ggttataaggat acattgagca actgactgaa atgcctcaaa atgttctt  
2521 cgatgcattt gggatatac aacgggtgt tatccagtga ttttttctc catttttagct  
2581 tccttagctc ctgaaagatc cataacttcg tatagcatac attatacga gttatgcggc

2641 cgcgacgtcc acatataacct gccgttcaact attatttagt gaaatgagat attatgatat  
 2701 tttctgaatt gtgattaaaa aggcaacttt atgcccattgc aacagaaaact ataaaaaata  
 2761 cagagaatga aaagaacacag atagatttt tagttctta ggcccgtagt ctgcaatcc  
 2821 ttttatgatt ttctatcaa caaaagagga aaatagacca gttcaatcc aaacgagat  
 2881 ctaatagaat gaggtcgaaa agtaaatcgc gcgggttgt tactgataaa gcaggcaaga  
 2941 cctaaaatgt gtaaagggc aagtgtatac tttggcgtca ccccttacat attttagtc  
 3001 ttttttatt gtgcgttaact aacttgccat cttcaaacag gagggctgga agaagcagac  
 3061 cgctaacaaca gtacataaaaa aaggagacat gaacgatgaa catcaaaaag tttcaaaaac  
 3121 aagcaacagt attaacctt actaccgcac tgctggcagg aggcgcactt caagcgtttg  
 3181 cgaaagaaaac gaaccaaaag ccatataagg aaacatacgg catttcccat attacacgcc  
 3241 atgatatgtc gcaaattccct gaacagcaaa aaaatgaaaa atatcaagtt cctgaatgtc  
 3301 attcgtccac aattaaaaat atctcttctg caaaaggcct ggacggttgg gacagctggc  
 3361 cattacaaaa cgctgacggc actgtcgcaa actatcacgg ctaccacatc gtcttgcatt  
 3421 tagccggaga tcctaaaaat gccgatgaca catcgattt catgttctat caaaatgtc  
 3481 gcgaaacttc tattgacagc tgaaaaaagc ctggccgcgt ctttaagac agcgacaaat  
 3541 tcgatgcaaa tgattctatc ctaaaagacc aaacacaaga atggtcaggt tcagccacat  
 3601 ttacatctga cggaaaaatc cggttattct acactgattt ctccggtaaa cattacggca  
 3661 aacaaacact gacaactgca caagttAACAGCAGCTCT ttgaacatca  
 3721 acggtgtaga ggattataaa tcaatcttgc acgggtacgg aaaaacgtat caaaatgtac  
 3781 agcagttcat cgatgaaggc aactacagct caggcgacaa ccatacgctg agagatcctc  
 3841 actacgtaga agataaaggc cacaataact tagtatttgc agcaaacact ggaactgaag  
 3901 atggctacca aggcaagaaa tctttatcca acaaagcata ctatggcaaa agcacatcat  
 3961 tcttcgtca agaaagtcaa aaacttctgc aaagcgatca aaaaacgcacg gctgatgttag  
 4021 caaacggcgc tctcggtatg attgagctaa acgatgatta cacactgaaa aaagtgtatga  
 4081 aaccgctgat tgcattcaac acagtaacag atgaaatttgc acgcgcgaaac gtctttaaaa  
 4141 tgaacggcaa atggtaacttgc ttcaactgact cccgcggatc aaaaatgacg attgacggca  
 4201 ttacgtctaa cgatatttac atgcttgggtt atgtttctaa ttctttaact ggcccataca  
 4261 agccgctgaa caaaactggc ctgtgtttaaa aatggatct tgatcctaaac gatgttaccc  
 4321 ttacttactc acacttgcgt gtacctcaag cggaaaggaaa caatgtcgatc attacaagct  
 4381 atatgacaaa cagaggatttgc tacgcagaca aacaatcaac gtttgcgcct agcttctgc  
 4441 tgaacatcaa aggcaagaaa acatctgttgc tcaaagacag catccttggaa caaggacaat  
 4501 taacagttaa caaataaaaaa cggaaaagaa aatgcccata tcctattggc attgacgtca  
 4561 ggtggcactt ttccggggaaa tttggcgggaa accccttattt gtttattttt ctaatacat  
 4621 tcaaataatgt atccgcctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa  
 4681 aggaagagta tgagtattca acatttccgt gtcgcctta ttccctttt tgccgattt  
 4741 tgccttcctg ttttgcattca cccagaaaacgc ctggtaaaatg taaatggatgc tgaagatcag  
 4801 ttgggtgcac gagggttgcattca acacttgc acgacttta aagttctgc atgtggcgc  
 4861 ttccgccttgc aagaacgttt tccaatgtatc acacttta aagttctgc atgtggcgc  
 4921 gtattatccc gtattgacgc cggcaagag caactcggtc gcccataca ctattctc  
 4981 aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta  
 5041 agagaattat gcagtgtgc cataaccatg agtgataaca ctgcggccaa cttaattctg  
 5101 acaacgatcg gaggaccgaa ggagctaac gctttttgc acaacatggg ggatcatgt  
 5161 actcgcccttgc atcggtggaa accggagctg aatgaagcca taccaaacga cgagcgtgac  
 5221 accacgatgc ctgttagcaat ggcacaaacgc ttgcgcaaaatc tattactgg cgaactactt  
 5281 actctagctt cccggcaaca attaatacgatc tggatggagg cggataaaatg tgcaggacca  
 5341 cttctgcgtc cggcccttcc ggctggctgg tttattgtcataaattctgg agccgggtgag  
 5401 cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta  
 5461 gttatctaca cggacggggatc tggcaact atggatgaaatc gaaatagaca gatcgctgag  
 5521 atagggtgcct cactgattaa gcatggtaa ctgtcagacc aagtttactc atatataactt  
 5581 tagattgatt taaaacttca tttttaattt aaaaaggatct aggtgaagat ctttttgc  
 5641 aatctcatga cccaaatccc ttaacgttag ttttcgttcc actgagcgtc agaccccgta  
 5701 gaaaagatca aaggatcttgc ttggatgc ttttttgc gcttaatctg ctgcttgc  
 5761 acaaaaaaac caccgctacc agcggtgggtt tttttgc gcttaatctg ctgcttgc  
 5821 ttccgaaagg taactggctt cagcagacgc cagataccaa atactgttct tcttagtgc  
 5881 ccgttagtag gccaccactt caagaactct gtagcaccgc ctacataccct cgctctgc  
 5941 atccgttac cagtggtgc tgccagtggc gataagtcgt gtcttaccgg gttggactca

6001 agacgatagt taccggataa ggcgccagcgg tcgggctgaa cgggggggttc gtgcacacag  
6061 cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagctgta gctatgagaa  
6121 agcgccacgc ttccccgaagg gagaaaaggcg gacaggtatc cggttaagcgg cagggtcggg  
6181 acaggagagc gcacgaggga gttccaggg gaaaacgcct ggtatctta tagtcctgtc  
6241 gggtttcgcc acctctgact tgagcgtcga tttttgtat gctcgtcagg ggggcccggc  
6301 ctatggaaaa acgcoagcaa cgcggccccc ttacggttcc tggccttttgc ctggcctttt  
6361 gctcacatgt tctttctgc gttatcccct gattctgtgg ataaccgtat taccgcctta  
6421 cgcgttaaaa acgacggcca gtagatctgt aatacgactc actatagggc gctagctgct  
6481 cgccgcagcc gaacgaccga gcgcagcggag tcagtgagcg aggaa (SEQ ID NO:03)

10

D: pLPS-Luc-EGFP

	2581	gcatgatgaa	cctgaatcgc	cagcggcattc	agcacccgtt	cgcccttcgt	ataatatttg
	2641	cccatggta	aaacggggc	gaagaagttt	tccatattgg	ccacgtttaa	atcaaaaactg
5	2701	gtgaaactca	cccaggatt	ggctgagac	aaaaacatat	tctcaataaa	ccctttaggg
	2761	aaataggcca	ggttttacc	gtAACACGCC	acatcttcg	aatatatgt	tagaaactgc
	2821	cgaaatcgt	cgtggattt	actccagac	gatgaaaac	tttcagttt	ctcatggaaa
10	2881	acgggttaac	aagggtgaac	actatccc	atcaccag	caccgtt	cattgccata
	2941	cggaaattccg	gatgacatt	catcaggcgg	gcaagaatgt	gaataaaaggc	cggataaaaac
	3001	ttgtgcttat	ttttcttac	ggtctttaa	aaggccgtaa	tatccagct	aacggctgg
	3061	ttataggtac	attgagcaac	tgactgaaat	gcctcaaaat	tttcttacg	atgcatttgg
15	3121	gatatatcaa	cggtggata	tccagtgatt	ttttctcca	tttagcttc	cttagctcct
	3181	gaaagatcca	taacttgc	tagcatacat	tatacgaagt	tatagatcca	atattattga
	3241	agatattatc	agggttattt	tctcatgac	ggatacatat	ttgaatgtat	ttagaaaaat
	3301	aaacaaatag	gggttcccg	cacatttccc	cgaaaagtgc	cacctgacgt	ggatctcgag
20	3361	ctcaagcttc	gaattcagg	tttcttgc	aatatcatac	ttatcctgtc	ccttttttt
	3421	ccacagctac	cggtcgcg	caagggcgag	gagctgttca	ccgggggtgg	gccccatctg
	3481	gtcgagctgg	acggcgacgt	aaacggccac	aagttcagcg	tgtccggcg	gggcgagggc
	3541	gatgccacct	acggcaagct	gaccctgaag	ttcatctgc	ccaccggcaa	gctgcccgtg
	3601	ccctggccca	ccctcg	gaccctgacc	tacggcgtgc	agtgttca	ccgctacccc
25	3661	gaccacatga	agcagcacga	cttcttcaag	tccgcacatgc	ccgaaggcta	cgtccaggag
	3721	cgcaccatct	tcttcaagga	cgacggcaac	tacaagaccc	gcccggaggt	gaagttcgag
	3781	ggcgacaccc	tggtaacc	catcgagct	aaggcatcg	acttcaagg	ggacggcaac
	3841	atccctggggc	acaagctga	gtacaactac	aacagccaca	acgtctata	catggccgac
	3901	aagcagaaga	acggcatcaa	ggtaactt	aagatccg	acaacatcg	ggacggcagc
	3961	gtgcagctcg	ccgaccacta	ccagcagaac	accccatcg	gacggccc	cgtgtcgctg
30	4021	cccgacaacc	actacctgag	cacccagtcc	gccctgagca	aagacccaa	cgagaagcgc
	4081	gatcacatgg	tcctgttgc	gttcgttgc	gccgcggg	tcactctcg	catggacgag
	4141	ctgtacaagt	aaagcgccg	cgactctaga	tcataatcg	ccataccaca	ttttagagg
	4201	tttacttgc	tttaaaaac	ctcccacacc	tcccccgt	cctgaaacat	aaaatgaatg
	4261	caattgttgc	tgttaacttgc	tttattgcag	cttataatgg	ttacaaataa	agcaatagca
	4321	tcacaaaattt	cacaaataaa	gcattttt	cactgcattc	tagttgtgg	ttgtccaaac
	4381	tcatcaatgt	atcttaaggc	gtaaattgt	agcgtaata	ttttgttaaa	attcgctt
	4441	aatttttgtt	aatcagctc	atttttaa	caataggcc	aatcggcaa	aatccctt
	4501	aaatcaaaag	aatagaccga	gataggggtt	agtgttgc	cagtggaa	caagagtcca
35	4561	ctattaaaga	acgtggactc	caacgtcaaa	ggcgaaaaaa	ccgtctatca	gggcgatggc
	4621	ccactacgt	aaccatcacc	ctaatacg	ttttgggt	cgaggtgccc	taaagacta
	4681	aatcggacc	ctaaagg	ccccgattt	agagcttgc	ggggaaagcc	ggcgaacgtg
	4741	gcgagaaagg	aaggaaagaa	agcgaaagg	gccccggct	ggcgttggc	aagtgttagcg
	4801	gtcacgctgc	gcgttacc	caccccgcc	gctttaatg	cgccgctaca	ggcgcgtca
40	4861	ggtggcactt	ttcgggaaa	tgtgcgcgg	acccctattt	gttattttt	ctaaatacat
	4921	tcaaatatgt	atccgc	catgacaataa	ccctgataaa	tgcttcata	atattgaaa
	4981	aggaagagtc	ctgaggcg	aagaaccac	tgtggatgt	gtgtcagtt	gggtgtggaa
	5041	agtccccagg	ctccccagg	ggcagaagta	tgcggatcat	gcatctca	tagtcagca
	5101	ccaggtgtgg	aaagtcccc	ggctcccc	caggcagaag	tatgcaaagc	atgcacatca
45	5161	attagtccgc	aaccatagtc	ccgcccc	ctccgc	cccgcccc	actccccc
	5221	gttccggccca	tttccccc	catggctgc	taattttt	tattatgc	gaggccgagg
	5281	ccgcctcgcc	ctctgagct	ttccaga	agtggagg	ctttttgg	ggcctaggct
	5341	tttgccaaaga	tgcataaga	gacaggat	ggatgtt	gcatgatt	acaagatgg
	5401	ttgcacgcag	gttctccgg	cgcttgggt	gagaggct	tccgctat	ctgggcacaa
50	5461	cagacaatcg	gctgtctg	tgccggcgt	ttccggct	cacgcgcagg	gcccgggtt
	5521	ctttttgtca	agaccgac	gtccgg	ctgaatg	tgcggat	ggcagcgccg
	5581	ctatcg	tgccac	ggcg	tgcgc	tgctcgac	tgtcactgaa
	5641	gccccggagg	actggctg	attggcgaa	gtgcgggg	agatctcc	gtcatctcac
	5701	cttgc	ccgagaa	atccatcat	gctgat	tgcggcgg	gcatacgctt
55	5761	gatccggcta	cctgcccatt	cgaccac	gcaaaatc	gcatcgac	agcacgtact
	5821	cgatggaaag	ccgttcttgc	cgatcagg	gatctgg	gacagcatc	ggggctcg
	5881	ccagccgaac	tgttgc	gctcaagg	agcat	ccggcgagga	tctcg

5941 acccatggcg atgcctgctt gccgaatatac atggtgaaa atggccgctt ttctggattc  
 6001 atcgactgtg gccggctggg tttggcggac cgctatcagg acatagcggt ggctaccgt  
 6061 gatattgtcg aagagcttgg cggcgaatgg gctgaccgct tcctcggtct ttacggatc  
 6121 gccgctcccg attcgacgcg catcgcccttc tatcgcccttc ttgacgagtt cttctgacg  
 6181 ggactctggg gttcgaaaatg accgaccaag cgacgcccgg cctgccccatca cgagatttc  
 6241 attccaccgcg cgcccttat gaaagggttgg gcttcggaaat cgttttccgg gacgcccgg  
 6301 ggtatgtatcc ccagcgcggg gatctcatgc tggaggttttt cgccccaccctt agggggaggc  
 6361 taactgaaac acggaaggag acaataccgg aaggaacccg cgctatgacg gcaataaaaaa  
 6421 gacagaataa aacgcacggg gtggggcgtt ttgttcataa acgcgggggtt cggtcccagg  
 6481 gctggcactc tgtcgatacc ccaccgagac cccattgggg ccaatacgcc cgcggttctt  
 6541 cctttcccccc accccacccc ccaagttcgg gtgaaggccc agggctcgcg gccaacgtcg  
 6601 gggcggcagg ccctgocata gcctcagggtt actcatatat acttttagatt gattnaaaaac  
 6661 ttcatttta attttaaaagg atcttaggtga agatccttt tgataatctc atgaccaaaaa  
 6721 tcccttaacg tgagtttcg ttccactgag cgtagaccc cgtagaaaag atcaaaggat  
 6781 cttcttgaga tcctttttt ctgcgcgtaa tctgctgctt gcaaaacaaaa aaaccaccgc  
 6841 taccagcggt gggttgggg cggatcaag agtaccaac tcttttccg aaggtaactg  
 6901 gcttcagcag agcgcagata ccaaataactg tccttctagt gtgcgttag tttaggcacc  
 6961 acttcaagaa ctctgttagca ccgcctacat acctcgctct gctaattctg ttaccagtgg  
 7021 ctgctgccag tggcgataag tcgtgtctta ccgggttggg ctcaagacga tagttaccgg  
 7081 ataaggcgca gcggtcgggc tgaacgggggg gttcggtcac acagcccagc ttggagcgaa  
 7141 cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc acgcttcccg  
 7201 aaggagaaaa ggcggacagg tatccggtaa gggcagggt cggAACAGGA gagcgcacga  
 7261 gggagcttcc agggggaaac gcctggatcc ttatagttc tgccgggtt cgccacctct  
 7321 gacttgagcg tcgattttt tgatgtctgt cagggggcgc ggcctatgg aaaaacgcca  
 7381 gcaacgcggc cttttacgg ttccctggct tttgctggcc tttgctcac atgttcttc  
 7441 ctgcgttatac ccctgattct gtggataacc gtattaccgc catgcat (SEQ ID NO:04)

### Example 3. Representative Splice Donor and Acceptor Sites

#### 30 A. Consensus Splice Donor and Acceptor oligos:

Consensus splice donor:  
(cloned into pDNR-1 at Apal and AvrII sites)

35 Site of Exon/intron boundary      |  
 top : CAGGTGAGTTAGGTAAGTGAACATGGTCATAGCTGTTTC  
 bottom: CCGGGTCCACTCAATCCATTCACTTGACCAGTATCGACAAAGGATC  
 (SEQ ID NOS: 05 & 06)

40 Consensus splice acceptor (includes branch site):  
(cloned into pEGFP-N1 at EcoRI and AgeI sites)

Site of Exon/intron boundary      |  
 top : AATTCAAGGGTTCCCTTGACAATATCATACTTATCCTGTCCCCTTTTTCCACAGCTA  
 bottom: GTCCCCAAAGGAACCTGTTATAGTATGAATAGGACAGGGAAAAAAAGGTGTCGATGGCC

(SEQ ID NOS:07 & 08)

#### 50 B. Splice donor from Human hemoglobin Beta

B, F & F Ref: CLON-069

Clontech Ref: P-90

F:\DOCUMENT\CLON\069\patent application.doc

Sequence encoding exon and intron sequence flanking the start of Human Hemoglobin Beta intron I:

5 Site of Exon/intron boundary |  
top : AGTTGGTGGTGAGGCCCTGGCAGGTTGGTATCAAGGTTACAAGACAGGT  
bottom: TCAACCACCACTCCGGGACCGTCCAACCATAAGTCCAATGTTCTGTCCA

(SEQ ID NOS: 09 & 10)

10 This splice donor sequence was encoded within the following oligo to enable cloning into pDNR-1 at the Apal and AvrII sites. Note that this oligo was additionally designed to place stop codons (TAG and TAA) in the two unused reading frames present in the MCS of pDNR-1. (The frame utilized is defined as starting with the first base of the loxP site in pDNR-1). In addition, remaining in 15 frame with the utilized frame is encoded an (HN)6 tag to enable protein purification in bacteria – this is encoded directly after the intron seq shown above.

20 Oligo for Splice Donor from Human Hemoglobin Intron I with added Stops and (HN)6 tag:

25 Site of Exon/intron boundary |  
Top : CGTAGTGTAAAGTTGGTGGTGAGGCCCTGGCAGGTTGGTATCAAGGTTACAAGACAGGTATAATCATAATCATAATCATAATCACAACTAGC  
Bottom: CCGGGCATCACATTCAACCACCTCCGGGACCGTCCAACCATAAGTCCAATGTTCTGTCCAGTATTAGTATTAGTATTAGTATTAGTATTAGTGTTC  
ATCGGATC

(SEQ ID NOS:11&12)

30 Sequence for (HN)6 tag within Splice donor oligo:

35 Top : GGT CAT AAT CAT AAT CAT AAT CAT AAT CAC AAC TAG  
Bottom: CCA GTA TTA GTA TTA GTA TTA GTA TTA GTG TTG ATC  
Peptide encoded: Gly His Asn His Asn His Asn His Asn His Asn stop

(SEQ ID NOS:13, 14 17 15)

40 Splice acceptor from Human hemoglobin Beta

This oligo encodes the splice Acceptor region of intron I from Human Hemoglobin Beta together with flanking exoon sequence. It was cloned into pEGFP-N1 at the 45 Agel and EcoR I sites.

Oligo for Human Hemoglobin Beta splice acceptor from Intron I:  
Site of Exon/intron boundary |

Top :  
AATTCTTGGGTTCTGATAGGCAGTGACTCTCTGCCATTGGCTATTTCCCACCCCTAGGCTGCTGGTGGTCTACC  
CTGGGACCTA  
Bottom:  
5 GAACCCAAAGACTATCCGTGACTGAGAGAGACGGCTAACAGATAAAAGGGTGGGAATCCGACGACCACAGATGGAAC  
CTGGGATGCC  
(SEQ ID NOS: 16 & 17)

It is evident from the above results and discussion that the subject  
10 invention provides an efficient method to transfer a nucleic acid from a first vector  
to a second vector, where the subject methods do not employ digestion and  
ligation protocols. Advantages provided by the subject invention include: the  
ability to transfer or clone a nucleic acid of interest from a single donor into a  
variety of different expression vectors at substantially the same time and in a  
15 known orientation and reading frame; the ability to readily identify successful  
clones; the ability to transfer many different genes to one or more expression  
vectors simultaneously; no longer needing to sequence the junctions of the  
transferred fragment and the expression vector or to resequence the gene  
transferred and the like. Another advantage of the subject invention is to provide  
20 for introns in the product vector, so as to remove any unwanted sequences from  
the final encoded product, and/or easily produce N- and/or C-terminal tagged  
fusion proteins. As such, the subject invention represents a significant  
contribution to the art.

25 All publications and patent applications cited in this specification are herein  
incorporated by reference as if each individual publication or patent application  
were specifically and individually indicated to be incorporated by reference. The  
citation of any publication is for its disclosure prior to the filing date and should not  
be construed as an admission that the present invention is not entitled to  
30 antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way  
of illustration and example for purposes of clarity of understanding, it is readily  
apparent to those of ordinary skill in the art in light of the teachings of this

invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.